Fernando H. Llano Alonso
Director

Joaquín Garrido Martín
Ramón Valdivia Jiménez
Coordinadores

INTELIGENCIA ARTIFICIAL Y FILOSOFÍA DEL DERECHO

Autores

Rafael de Asís Roig
Nuria Belloso Martín
Stefano Bini
Roger Campione
Thomas Casadei
Miguel de Asís Pulido
Daniel García San José
Joaquín Garrido Martín
Ana Garriga Domínguez
Mª Isabel González Tapia
Fernando H. Llano Alonso
Leonor Moral Soriano
Stefano Pietropaoli
Álvaro Sánchez Bravo
Adolfo J. Sánchez Hidalgo
Mª Olga Sánchez Martínez
María Sepúlveda Gómez
José Ignacio Solar Cayón
Ramón Valdivia Jiménez
Diana Carolina Wisner Glusko
Laura Gómez Abeja

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN
AGENCIA ESTATAL DE INVESTIGACIÓN

Laborum EDICIONES

FERNANDO H. LLANO ALONSO
Director

JOAQUÍN GARRIDO MARTÍN
RAMÓN VALDIVIA JIMÉNEZ
Coordinadores

INTELIGENCIA ARTIFICIAL Y FILOSOFÍA DEL DERECHO

Autores:

RAFAEL DE ASÍS ROIG
NURIA BELLOSO MARTÍN
STEFANO BINI
ROGER CAMPIONE
THOMAS CASADEI
MIGUEL DE ASÍS PULIDO
DANIEL GARCÍA SAN JOSÉ
JOAQUÍN GARRIDO MARTÍN
ANA GARRIGA DOMÍNGUEZ
Mª ISABEL GONZÁLEZ TAPIA
FERNANDO H. LLANO ALONSO
LEONOR MORAL SORIANO
STEFANO PIETROPAOLI
ÁLVARO SÁNCHEZ BRAVO
ADOLFO J. SÁNCHEZ HIDALGO
Mª OLGA SÁNCHEZ MARTÍNEZ
MARÍA SEPÚLVEDA GÓMEZ
JOSÉ IGNACIO SOLAR CAYÓN
RAMÓN VALDIVIA JIMÉNEZ
DIANA CAROLINA WISNER GLUSKO
LAURA GÓMEZ ABEJA

Laborum
EDICIONES

Esta publicación es parte del proyecto de I+D+i PID2019-108155RB-I00, financiado por MCIN/ AEI /10.13039/501100011033

Edita:
Ediciones Laborum, S.L.
Avda. Gutiérrez Mellado, 9 - Planta 3ª, Oficina 21 - 30008 Murcia
Tel.: 968 24 10 97
e-mail: laborum@laborum.es
www.laborum.es

ISBN edición digital: 978-84-19145-25-3
ISBN edición papel: 978-84-19145-21-5
Depósito Legal: MU 679-2022

A Antonio Enrique Pérez Luño,
maestro de iusfilósofos y pionero
de los estudios sobre Ciberderecho e
Informática jurídica en España

"El hombre actual no sabe qué ser, le falta imaginación para inventar el argumento de su propia vida"

(JOSÉ ORTEGA Y GASSET,
MEDITACIÓN DE LA TÉCNICA, 1939)

ÍNDICE

II. FILOSOFÍA Y ÉTICA DE LA INTELIGENCIA ARTIFICIAL

INTRODUCCIÓN

A primeros de diciembre del año 2021 se celebró, de forma presencial, en el Salón de Grados de la Facultad de Derecho de la Universidad de Sevilla, la segunda edición del *Congreso Internacional sobre Inteligencia Artificial, Robótica y Filosofía del Derecho*[1]. A diferencia de la mayoría del seminarios y congresos jurídicos que se habían organizado hasta entonces en torno a las implicaciones de la Inteligencia Artificial y las Nuevas Tecnologías en el mundo del Derecho contemporáneo, este congreso, realizado en el marco del Proyecto de Investigación PID2019-108155RB-I00/AEI/10.13039/501100011033 "Biomedicina, Inteligencia Artificial, Robótica y Derecho: los retos del jurista en la era digital", con la colaboración oficial de la Agencia Española de Investigación del Ministerio de Ciencia e Investigación, reunió a destacados especialistas en Nuevas Tecnologías, Informática, Robótica e Inteligencia Artificial jurídica dentro del panorama iusfilosófico español e italiano contemporáneo.

Durante la presentación de la conferencia inaugural, dictada por el Prof. Rafael de Asís Roig[2], recordé que a la Filosofía del Derecho le corresponde el honor de haber sido la primera disciplina jurídica en haberse ocupado de cuestiones relacionadas con el Ciberderecho, la Informática jurídica, las Nuevas Tecnologías y su impacto en el ámbito de los derechos y las libertades de los ciudadanos en la sociedad global de la información. En este sentido, los primeros estudios iusfilosóficos sobre esta materia específica se publicaron hace ya más de cincuenta años en Italia, Inglaterra y España. A este respecto, entre nuestros clásicos de referencia destacan los trabajos de Mario G. Losano: *Giuscibernetica. Macchine e modelli giuscibernetici nel diritto*, Einaudi, Torino, 1965; Vittorio Frosini: *Il diritto nella società tecnológica*, Giuffrè, Milano, 1981; Colin Tapper: *Computers and the Law*, Weidenfeld and Nicolson, London, 1973; Antonio E. Pérez Luño: *Cibernética, Informática y Derecho. Un análisis metodológico*, Publicaciones del Real Colegio de España, Bolonia, 1976.

La investigación sobre Inteligencia Artificial, Robótica e Informática Jurídica en clave iusfilosófica ha tenido su continuidad hasta el presente a

[1] La primera edición de este congreso internacional hubo de realizarse, debido a las restricciones de la pandemia COVID-19, a través de un webinar durante los días 9-11 de diciembre de 2020, y fue seguido regularmente por más de trescientas personas, lo cual da una idea aproximada del interés que suscitó el contenido del programa. A resultas de este primer congreso sobre Inteligencia Artificial y Derecho, se publicó el libro *Inteligencia Artificial y Derecho. El jurista ante los retos de la era digital,* Fernando H. Llano Alonso y Joaquín Garrido Martín (eds.), Cizur Menos (Navarra), Thomson Reuters Aranzadi, 2021.

[2] La conferencia del Prof. Rafael de Asís llevaba por título: "Derecho(s) y tecnologías convergentes".

través de los ensayos de, entre otros autores (y sin ánimo de exhaustividad), Enrico Pattaro, *Intelligenza Artificiale e diritto dell´ambiente*, Edizione di Documentazione del Consiglio Regionale dell´Emilia-Romagna, Bologna, 1991; Giovanni Sartor: *Artificial Intelligence in Law*, Tano, Oslo, 1993; Ugo Pagallo: *The Laws of Robots*. *Crimes, Contracts, and Torts*, Dordrecht-Heidelberg-New York-London, Springer, 2013; Rafael de Asís Roig: *Una mirada a la robótica desde los derechos humanos*, Madrid, Dykinson-Instituto de Derechos Humanos Bartolomé de Las Casas. Universidad Carlos III de Madrid, 2014; *Derecho y tecnologías*, Madrid, Dykinson-Universidad Carlos III de Madrid, 2022; José Ignacio Solar Cayón, *La Inteligencia Artificial Jurídica*, Cizur Menor (Navarra), Thomson Reuters Aranzadi, 2019; Thomas Casadei y Stefano Pietropaoli (eds.), *Diritto e tecnologie informatiche*, Milano, Wolters Kluwer, 2021; Fernando Llano Alonso y Joaquín Garrido Marín (eds.), *Inteligencia Artificial y Derecho. El jurista ante los retos de la era digital*, Cizur Menos (Navarra), Thomson Reuters Aranzadi, 2021; Luciano Floridi, *Etica dell´intelligenza artificiale. Sviluppi, opportunità, sfide*, Milano, Raffaello Cortina Editore, 2022.

El presente volumen reúne los trabajos elaborados en forma de capítulos por los ponentes que participaron en la segunda edición del ya mencionado Congreso Internacional sobre Inteligencia Artificial, Robótica y Filosofía del Derecho. Aunque la perspectiva jurídica desde la que se aborda el objeto principal de estudio de este libro es interdisciplinar, los veintiún capítulos que lo integran tienen un denominador común: todos se orientan al estudio de cuestiones que guardan relación directa con muchos de los temas centrales de la Teoría y la Filosofía del Derecho, la teoría de la justicia y los derechos humanos en la era digital.

La obra se estructura en cuatro partes bien diferenciadas, pero a la vez complementarias entre sí:

El primer bloque temático versa sobre Inteligencia Artificial, derechos y libertades, y en él se engloban los capítulos de Rafael de Asís Roig, a propósito de los retos éticos que plantean las tecnologías convergentes a la sociedad contemporánea, en particular a propósito de la protección de la identidad y la suficiencia del enfoque de derechos humanos; Nuria Belloso Martín, que aporta razones justificativas de la necesidad de desarrollar el Derecho de protección de los sesgos algorítmicos (con especial referencia a los que discriminan por razón de género); Stefano Bini, quien considera indispensable mantener la dimensión instrumental de la tecnología ante el dilema humanidad-artificialidad sobre todo en relación con la buena gestión y el funcionamiento correcto del sistema judicial contemporáneo; Laura Gómez Abeja, que analiza el impacto del *big data* y el uso de algoritmos (para la toma de decisiones que afectan a personas) concretamente en dos derechos fundamentales: el derecho a la protección de

datos personales (art. 18.4 CE) y el derecho a la no discriminación (art. 14 CE); y Mª Olga Sánchez Martínez, que se centra en el estudio de la posverdad en la sociedad de las nuevas tecnologías y de la comunicación que, paradójicamente, pueden ser utilizadas justamente para todo lo contrario, es decir, para generar desinformación y polarización en la sociedad si se hace un uso inadecuado de las mismas.

La segunda parte comprende varios temas de naturaleza filosófica y de ética de la Inteligencia Artificial de los que se ocupan en sus respectivos capítulos los siguientes autores: Thomas Casadei, que reflexiona a propósito de las transformaciones del Derecho en la era de la ciudadanía digital (tanto a raíz del impacto de la red y de las nuevas tecnologías en la experiencia jurídica, como por la transformación de las profesiones jurídicas y forenses), en el estudio en torno al fenómeno de las brechas digitales en la era de la ciudadanía digital, así como en la enseñanza y el estudio del Derecho en la era digital; Joaquín Garrido Martín, quien, por su parte, se aproxima, desde un punto de vista filosófico-científico, al estudio del impacto de la neurociencia, las ciencias cognitivas y computacionales, y el fenómeno del progresivo automatismo de la mente humana debido al desarrollo de la Inteligencia Artificial y la robótica avanzada; Fernando H. Llano Alonso, que se ocupa del problema de la deshumanización del mundo real y de la absorción del *homo faber* en el universo del metaverso, lo cual nos pone de relieve, cada vez de forma más patente, la enajenación de la identidad humana en el mundo de las no-cosas (parafraseando a Byung-Chul Han); Stefano Pietropaoli, el cual profundiza en el fenómeno de la deshumanización del Derecho y en la aparición en el escenario digital sin precedentes de nuevas personas no-biológicas *sui iuris* que lleguen a compartir (e incluso acaben arrebatando) la condición de personas a los seres humanos; por último, a propósito del futuro de la libertad religiosa ante el anunciado advenimiento de la singularidad tecnológica; y Ramón Valdivia Jiménez, que invoca la ética humanista de Hans Jonas y Jürgen Habermas para preguntarse por la oportunidad de recuperar unos principios ético-jurídicos que salvaguarden a la humanidad de una posible deriva excesivamente controladora de la IA en materia de religión.

La tercera parte contiene seis capítulos a propósito de la robótica y la Inteligencia Artificial jurídica: Roger Campione se adentra en el ámbito del *ius in bello* y las armas autónomas en la era de la revolución tecnológica 4.0, y pone de manifiesto cómo el uso de drones, los sistemas de armamento autónomo, (*Lethal Autonomous Weapon Systems* - LAWS), las formas de potenciamiento humano mediante la técnica, y los sistemas autónomos de inteligencia artificial que permiten reemplazar soldaos humanos por máquinas puede contribuir a reducir el coste de vidas humanas, pero también plantea otras hipótesis ciertamente inquietantes, por ejemplo, el supuesto de que se adquiriesen *killer*

robots por parte de regímenes represivos irrespetuosos con el Derecho internacional y desconsiderados hacia los derechos humanos; Miguel de Asís Pulido centra su estudio en torno a tres posibles usos de la Inteligencia Artificial aplicada al proceso: el empleo de la justicia predictiva para solucionar el colapso que sufre nuestra Administración de Justicia, la fiscalización de sentencias para detectar ciertos sesgos o patrones discriminatorios presentes en el sistema jurídico, y el pronóstico para definir la estrategia procesal de los abogados y sus representados; Mª Isabel González Tapia reflexiona en su capítulo sobre si el Derecho penal debe intervenir en el ámbito de los neuroderechos frente a los riesgos potenciales y el uso directo de las neurotecnologías sobre el ser humano y, en caso afirmativo, en qué términos debería hacerlo; Adolfo Sánchez Hidalgo nos acerca en su estudio a la cuestión de la personalidad electrónica de los robots teniendo en cuenta la complejidad de este objeto de estudio y las diferentes perspectivas desde las que se puede contemplar este asunto: una visión antropomórfica de estos sistemas de IA en clave religiosa, filosófica, neurológica y jurídica, por citar solamente algunos de los principales puntos de vista; el capítulo de María Sepúlveda Gómez se sitúa en el marco del Derecho del trabajo y de la Inteligencia Artificial, y en él se justifica la necesidad de realizar un proceso de transición digital del mercado de trabajo que sea justa, porque el nuevo paradigma técnico-económico no puede renunciar a ser también social, ahora bien, para llevar a cabo una transición verdaderamente justa, es preciso fortalecer (y no debitar) los valores y los principios propios del Estado social y democrático; por último, José Ignacio Solar Cayón se centra en uno de los temas de su especialidad: la Inteligencia Artificial jurídica y la justicia digital, y nos invita a conocer las principales herramientas de inteligencia artificial que se están empleando en diversas jurisdicciones y las experiencias internacionales más avanzadas en el diseño de tribunales en línea, a fin de ofrecernos una visión panorámica de las ventajas y los riesgos que comportan estas tecnologías.

El cuarto bloque abarca contiene cinco capítulos sobre temas vinculados al Derecho internacional y de la Unión Europea, el Estado de Derecho y la Administración digital: Daniel García San José estudia el significado y la justificación de los valores que inspiran la Carta de las Naciones Unidas en la regulación internacional de la Inteligencia Artificial, centrándose en particular en la Recomendación sobre la ética de la Inteligencia Artificial aprobada por la UNESCO el 23 de noviembre de 2021; Ana Garriga Domínguez reflexiona sobre la obtención de perfiles, particularmente ideológicos, y su relación con las libertades de expresión e información, las *fake news* y el fenómeno de la desinformación a través de la red; Leonor Moral Soriano pretende demostrar que los sistemas de decisión automatizada resultan aplicables al Derecho Administrativo, sobre todo cuando se utilizan en la adopción de actos

administrativos; sin embargo, para que la tecnología no acabe adueñándose del razonamiento jurídico, y el juez-robot no termine usurpando la posición del juez humano, es preciso que, al usar estos sistemas de decisión automatizada, se observen los principios del Derecho Administrativo y la normas relativas a la competencia, el procedimiento administrativo como garantía, y la motivación de los actos administrativos (esencial para su revisión); Álvaro Sánchez Bravo analiza en detalle el contenido de la Estrategia Nacional de Inteligencia Artificial (denominada España Digital 2025), que consta de casi medio centenar de medidas articuladas en torno a diez ejes estratégicos: conectividad digital, impulso a la tecnología 5G, competencias digitales, ciberseguridad, transformación digital del sector público, transformación digital de la empresa y emprendimiento digital, proyectos tractores de digitalización sectorial, España como *hub* audiovisual, economía del dato e Inteligencia Artificial, y derechos digitales; por último, Diana Carolina Wisner Glusko razona sobre la importancia del Estado de Derecho en el desarrollo del marco legal sobre los sistemas de inteligencia artificial en la Unión Europea e intenta responder a tres grandes cuestiones: en primer lugar, en qué medida las propuestas europeas para la regulación de la IA se fundamentan en la garantía del Estado de Derecho; en segundo lugar, si el derecho condiciona cómo se desarrollan y se aplican esas tecnologías vinculadas a la IA o si, por el contrario, son estas tecnologías disruptivas las que dan forma al derecho; y, finalmente, cómo debería ser la inteligencia artificial para no contribuir al debilitamiento del Estado del Derecho e inclusive para mantenerlo y reforzarlo.

En la pluralidad de los estudios que se unifican en este volumen bajo el rótulo de *Inteligencia Artificial y Filosofía del Derecho* se aportan algunas de las claves necesarias para entender, en toda su complejidad y con sentido crítico, el cambio de paradigma que se está operando en muchos de los conceptos, las categorías e instituciones del Derecho y la justicia, y también se proporcionan al lector argumentos válidos para responder a muchas de las interrogantes que, a propósito del impacto de la Inteligencia Artificial en la experiencia jurídica contemporánea, interpelan tanto al estudioso como al profesional del Derecho.

El objetivo principal que persigue esta obra es, precisamente, contribuir al debate doctrinal sobre los itinerarios que tendrán que seguir y los retos que habrán de afrontar los juristas ante el horizonte cada vez más cercano de la singularidad tecnológica. Confiamos en que el lector de este libro encuentre en él razones que justifiquen el cumplimiento de este propósito.

FERNANDO H. LLANO ALONSO
Catedrático de Filosofía del Derecho
Facultad de Derecho. Universidad de Sevilla

I. INTELIGENCIA ARTIFICIAL, DERECHOS Y LIBERTADES

CAPÍTULO I

ÉTICA, TECNOLOGÍA Y DERECHOS

RAFAEL DE ASÍS ROIG
Instituto de Derechos Humanos Gregorio Peces-Barba
Universidad Carlos III de Madrid
rafael.asis@uc3m.es

En los últimos años, la discusión sobre los retos éticos que plantean las tecnologías convergentes ha crecido significativamente. Jeroen Hopster (Hopster, 2021, 133 y ss.), ha planteado un marco ético de análisis de estas tecnologías dependiendo del tipo de tecnología, de su significado y de su proyección. En esos ámbitos es común aludir a problemas relacionados con la igualdad y no discriminación, la autonomía, la responsabilidad, la privacidad, la intimidad… Pues bien en este trabajo, me voy a referir a dos de estos grandes retos: la protección de la identidad y la suficiencia del enfoque de derechos humanos. Se trata de dos retos que esconden en su interior una problemática compleja.

En efecto, la protección de la identidad exige en primer lugar saber si es posible hablar de una identidad humana o de una identidad personal y, además, cómo protegerla sin entrar en colisión con la diversidad y la no discriminación. Por su parte, la discusión sobre la suficiencia del enfoque de derechos humanos, nos dirige a la reflexión sobre si el tratamiento de estas tecnologías requieren de algo más que normas sobre derechos humanos y si es necesario el reconocimiento de nuevos derechos.

La complejidad en si de estas dos cuestiones se multiplica al proyectarse en el ámbito de unas tecnologías que están en continua evolución y parecen ser mucho menos regulables que otros fenómenos.

1. LA PROTECCIÓN DE LA IDENTIDAD

La cuestión de la identidad es un problema importante en el campo de los derechos humanos porque es aquello que está detrás de su principal fundamento: la dignidad humana. Ésta, parte de una idea de ser humano compuesta por una serie de atributos que lo singularizan y que se expresan en términos de identidad. De esta forma, el análisis de la justificación de los derechos irremediablemente lleva a esta cuestión. Pero también, el interés por la identidad puede surgir si consideramos que se trata precisamente de una idea que está en la base de la discriminación de grupos de personas y, de manera especial, de las personas con discapacidad (De Asís, 2013). Desde esta segunda dimensión adquiere relevancia lo que podemos denominar como identidad personal. Y ello porque esta idea permite subrayar lo singular de cada uno, la

atención a la persona, la diversidad y la dimensión personal de la vida humana digna.

Por otro lado, el problema de la alteración de la identidad, es una cuestión presente en muchas de las llamadas tecnologías convergentes. Lo está en la genética, principalmente en la actualidad a través del uso del método de edición genética CRISPR, que permite el corte y la edición con alta precisión de información genética en el ADN de cualquier organismo vivo, incluidos los humanos. También está presente en el campo de la Inteligencia Artificial, principalmente con su repercusión en la esfera privada y la intimidad a través de las técnicas de Big Data. Lo está en la Robótica de la mano de la reflexión sobre el concepto de robot y a través de la presencia cada vez mayor del fenómeno de los ciborgs. Y claramente lo está también en el campo de la neurotecnología, en donde se ha llegado a afirmar que la cuestión de la identidad es la que singulariza a la llamada neuroética frente a otras éticas aplicadas (De Asís, 2022, 54).

1.1. Identidad humana e identidad personal

Analizar cómo afectan las tecnologías a la identidad requiere, en primer lugar, tener claro un concepto de ésta. Sin embargo, el concepto de identidad es un concepto controvertido.

Es posible referirse así a diferentes visiones de la identidad: la biológica, que se centra en rasgos de ese tipo que permanecen invariables; la narrativa para la que la identidad tiene que ver con el relato de la vida; la relacional en la que la identidad tiene que ver con el reconocimiento de otros: la psicológica… (Wiggins, 1980; Shoemaker, 1996).

La identidad es, en ocasiones, concebida como naturaleza humana. Pero esta identificación puede ser problemática. Así por ejemplo, la Real Academia de la Lengua define la naturaleza humana como: "cualidad de los seres humanos no modificada por la educación". Sin embargo, probablemente, muchos estemos de acuerdo en considerar que la identidad tiene que ver también con el contexto.

Tradicionalmente, la identidad se ha identificado a través de dualidades (Ferrater Mora, 1965a, 402; Ferrater Mora, 1965b, 903): cuerpo y alma en la tradición cristiana (Agustín de Hipona, 2002); cuerpo y mente (Descartes, 1989, 178 y 179); más recientemente, biología y convención (Mosterín, 2013, 103 y ss.)… Ahora bien, ¿es lo mismo identidad humana que identidad personal? Volviendo al Diccionario de la Real Academia, es posible observar cómo, al referirse a la identidad, nos proporciona una definición de la identidad personal.

En efecto entre las diferentes acepciones del término identidad presentes en el diccionario, hay dos que nos interesan para nuestra reflexión. En virtud de la primera, identidad es el conjunto de rasgos propios de un individuo o de una colectividad que los caracterizan frente a los demás. En virtud de la segunda, se trata de la conciencia que una persona o colectividad tiene de ser ella misma y distinta a las demás.

Los rasgos propios van a ser diferentes dependiendo de la visión de la identidad que se maneje. Desde una visión biológica podemos en este punto hablar de órganos, pero a los órganos deberíamos añadir procesos como el pensar, sentir y actuar… Por su parte, el diccionario al hablar de conciencia de una misma, de la autoconciencia, parece estar haciendo referencia al conocimiento y por lo tanto al ejercicio de facultades como la percepción, la voluntad, la imaginación, la memoria, la intuición, la razón…

Ahora bien, una concepción de la identidad personal basada exclusivamente en lo anterior, puede parecer insuficiente. Gregorio Peces-Barba, al referirse al marco justificativo de los derechos humanos, señalaba cómo éste partía de la libertad psicológica entendida como el libre albedrío y se dirigía hacia el logro de la libertad moral entendida como la satisfacción de los planes de vida (Peces-Barba, 1989, 265 y ss.). No cabe duda de que libre albedrío y plan de vida, son también componentes de la identidad personal.

Pero además, parece que una concepción de la identidad personal que no tiene en cuenta el contexto y la experiencia es insuficiente. La identidad personal se confecciona también desde estos dos referentes que, en cierta forma, tienen que ver con otro de los rasgos que se predican de lo humano, su sociabilidad.

Así, desde todo lo anterior, es posible construir una idea de identidad personal basada en dos dimensiones, la condición y la situación (Barranco Avilés, 2016). Estas dos dimensiones se utilizan por ejemplo en el ámbito de la identificación biométrica, en el que se analizan características físicas pero también características de comportamiento[1]. Condición y situación son dos elementos clave en la consideración de las personas, como lo demuestra la

[1] La propuesta de Reglamento de la Unión Europea sobre Inteligencia Artificial, de 21 de abril de 2021, define los datos biométricos como: "datos personales resultantes de un tratamiento técnico específico relacionado con las características físicas, fisiológicas o de comportamiento de una persona física, que permiten o confirman la identificación única de esa persona física, como imágenes faciales o datos dactiloscópicos". En este ámbito, se hacen mediciones que pueden ser fisiológicas (huellas dactilares, patrón de las venas, iris y retina del ojo, forma de la cara, ADN, sangre, saliva, orina) o de comportamiento (reconocimiento de voz, dinámica de la firma, dinámica de la pulsación de las teclas, sonido de los pasos, gestos…).

definición actual de persona con discapacidad que combina precisamente estos dos elementos entendidos como deficiencia (condición) y barreras (situación). Dentro de este campo es significativo como en muchos ordenamientos jurídicos que están adaptando su legislación en materia de capacidad jurídica a la Convención sobre los Derechos de las Personas con Discapacidad de Naciones Unidas, como es el caso de España, la identidad cobra un especial relieve, por ejemplo a través de la determinación de la trayectoria vital de la persona, referencia obligada a la hora de llevar a cabo, de manera excepcional, funciones representativas.

Ahora bien, ¿donde radica la distinción entre identidad humana e identidad personal? La identidad humana puede presentarse como una suerte de universalización de las identidades personales. Por eso, en su comprensión, normalmente se deja un lado la dimensión de la situación al entenderse que no es universalizable. Esto hace que su uso sea problemático ante la diversidad de lo humano (aunque la condición tampoco sea completamente universal). Por su parte, la identidad personal individualiza la condición de la identidad humana (la condición humana) y expresa su diversidad. Además, subraya la dimensión de la situación, permitiendo entender y descubrir las barreras que afronta una persona.

Como señalé antes, identidad humana e identidad personal, son dos ideas necesarias en la fundamentación de los derechos. La dignidad humana se basa en una idea de identidad humana, y se satisface cuando es posible hablar de una vida humana digna, esto es, cuando se respeta la identidad personal. Y es que, como ha señalado J.L. Piñar Mañas "en gran medida la historia ha oscilado entre los intentos del poder por controlar, definir y tergiversar la identidad de las personas y la lucha del ser humano por alcanzar la propia identidad" (Piñar Mañas, 2018, 98). Así, de alguna manera, la identidad humana forma parte del metafundamento de los derechos mientras que la identidad personal constituye el fundamento (De Asís, 2001).

El ataque a la identidad personal puede venir básicamente de dos maneras. Puede surgir a través de la realización de daños a los órganos y funciones que están en la condición, pero también a través de la proliferación de barreras de todo tipo (como por ejemplo la institución de la incapacitación jurídica). Por su parte, la singularidad del ataque a la identidad humana (que siempre será también un ataque a la identidad personal) tiene que ver con la posibilidad de realizar un daño a aquellos órganos que integran la condición, un daño que provoque que no se pueda configurar la identidad o que esta sea modificada y alterada.

1.2. Tecnologías e identidad

Como ya he señalado, las tecnologías convergentes pueden afectar tanto a la identidad humana como a la identidad personal. En lo que sigue abordaré dos ejemplos de ello: el de la diversidad, que se proyecta principalmente en la identidad personal, y el de la mejora, que se proyecta principalmente en la identidad humana (y por extensión en la personal).

1.2.1. La diversidad

La idea de diversidad ha venido adquiriendo una importancia singular dentro del discurso de los derechos humanos. Sin embargo, diversidad y tecnologías convergentes tienen una relación problemática. Muchas de estas tecnologías, se apoyan en grandes cantidades de datos desde los que se realiza una suerte de ejercicio de generalización. Y en este ejercicio de generalización, la diversidad se ve amenazada principalmente por dos razones supuestamente enfrentadas y que podemos describir como la atención a la diversidad y la desatención a la diversidad.

Esta última, la desatención a la diversidad, produce en muchos casos situaciones de discriminación que se hacen muy evidentes en el marco del acceso a la tecnología. Pero está también presente en uno de los problemas más tratados cuando se analiza la repercusión social de la tecnología: los sesgos discriminatorios. Estos son consecuencia de la existencia de actitudes y visiones discriminatorias que pasan inadvertidas en las aplicaciones. Por eso, la lucha contra los sesgos requiere, en primer lugar, transparencia y publicidad; en segundo lugar, participación; y en tercer lugar, formación. Y todos estos planos requieren atender a la diversidad.

Sin embargo, la atención a la diversidad puede ser también una amenaza para la propia diversidad. Un buen ejemplo de ello lo constituye el enfoque con el que se suele abordar la discapacidad desde la tecnología, a través de proyectos e instrumentos cuyo objetivo es acabar con la discapacidad, en muchos casos desde el presupuesto de que es una enfermedad que puede ser curada gracias a la tecnología. La tecnología se presenta como oportunidad para hacer que las personas con discapacidad se sitúen en las mismas condiciones que el resto de personas (Hall, 2020). El enfoque mayoritario de la discapacidad desde el punto de vista tecnológico es un enfoque que se centra más en el aspecto de la condición que en el de la situación. De alguna manera pretende acabar con la diversidad de la condición, considerándola como un problema, sin atender aquello que más discrimina a las personas con discapacidad: su situación. La protección de la diversidad requiere, principalmente, luchar contra las barreras que colocan a las personas con discapacidad en una situación de discriminación y no tanto luchar contra las llamadas deficiencias.

Precisamente por eso, el discurso de los derechos aborda la cuestión de la vulnerabilidad principalmente como parte de la dimensión social de la persona y no tanto, que también, como parte de la dimensión ontológica.

1.2.2. La mejora: especial referencia a la mejora moral

He destacado antes como una de las consecuencias de las aplicaciones tecnológicas puede ser la producción de un daño a la identidad. Ese daño puede tener tal magnitud que provoque que ésta no pueda configurarse o se vea modificada y alterada (Llano, 2018, 122). Este peligro está muy presente con el uso de ciertas aplicaciones neurotecnológicas, como se han encargado de enfatizar los defensores de las propuestas de neuroderechos.

En cualquier caso, conviene llamar la atención sobre cómo en muchos casos, las propuestas de neuroderechos que se presentan dentro de un discurso que parte de los peligros presentes en ciertos usos de las neurotecnologías sobre el cerebro, en realidad terminan defendiendo este uso para la mejora. De hecho, los dos autores que se refieren por primera vez a la libertad cognitiva, Boire y Sententia, que como es sabido es uno de los derechos presentes en las distintas propuestas de neuroderechos, son abiertos defensores de esta mejora (Boire, 1999, 7; Sententia, 2004, 221).

Pues bien, en este punto no me voy a centrar así en el daño sobre la identidad sino en su mejora. Y es que, efectivamente, las tecnologías se utilizan para la mejora humana, lo que sin duda tiene consecuencias sobre la identidad.

1.3. Identidad y mejora: la mejora moral

La cuestión de la mejora puede considerarse como un tema clásico en el tratamiento de la ciencia y la tecnología desde un punto de vista ético. A pesar de ello no se trata de un tema resuelto ya que en su tratamiento hay que hacer frente a tres cuestiones distas de ser claras (Lema, 2015, 367 y ss.). Por un lado la distinción entre terapia y mejora; por otro la dificultad de plantear la existencia de mejoras absolutas (esto es, consideradas así por todos); y, por último, de nuevo, la propia noción de identidad humana (necesaria para saber, precisamente, cuando se produce una mejora).

En todo caso es un problema abordado no solo desde la reflexión ética sino también en el campo del Derecho, en el que es habitual rechazar intervenciones de mejora, por ejemplo, en el campo de la genética.

La justificación de la mejora es uno de los rasgos que caracterizan el pensamiento transhumanista (Llano Alonso, 2018, 26). Así la tensión terapia-mejora es el eje sobre el que gira la conocida fábula de Nick Bostrom del dragón tirano (Bostrom, 2005, 273 y ss). El envejecimiento humano es un dragón tirano que devora miles de vidas cada día y que es admitido por nuestra

sociedad invirtiendo muchos recursos en su mantenimiento. Pero la tecnología está planteando la posibilidad de acabar con el dragón, de "curar" el envejecimiento. Para ello necesita también de recursos cuya administración no hay que demorar…

Dos son los argumentos principales que se suelen esgrimir para rechazar la mejora. Por un lado, está el argumento de la discriminación, que tiene que ver tanto con el acceso a la intervención como con sus consecuencias (Nussbaum, 2002, 6 y 7). Por otro, está el argumento del daño a la identidad humana, incluso cuando la intervención es consentida (Sandel, 2007, 127). Así, se ha señalado que no son admisibles las intervenciones de mejora consentida cuando pueden producir un daño o un cambio en la identidad humana.

Sin embargo, el pensamiento transhumanista ha contestado a este último argumento a través de una propuesta que tensa aún más esta reflexión y que es la de la mejora moral. El argumento es muy sencillo: si conseguimos hacer a las personas mejores desde un punto de vista moral, la intervención estaría justificada.

En este sentido, en 2008, Tom Douglas publica su trabajo "Moral enhancement" pretendiendo demostrar que es posible llevar a cabo intervenciones de mejora favorables para la humanidad en general (Douglas, 2008). Según Douglas, conseguir mejorar las motivaciones morales de las personas debería estar permitido.

También en 2008, Ingmar Persson y Julian Savulescu, se refieren a la mejora moral como herramienta para solventar los posibles males que conllevaría la mejora cognitiva (Persson/Savulescu, 2008). La cuestión de la mejora moral de las personas es una cuestión que ha estado siempre presente en la historia de la humanidad. Herramientas como el Derecho y, sobre todo, la educación se ha justificado en ocasiones desde esta referencia. Así, según estos autores, estas dos herramientas de control social, la educación y el Derecho, son insuficientes para conseguir que los seres humanos actúen por el bien de todos. Para ellos, la mejora del comportamiento moral no sólo es éticamente aceptable sino que es algo exigible (Lara Sánchez, 2016).

Entre las distintas críticas lanzadas a las propuestas de mejora moral, más allá de aquellas que tienen que ver con su posibilidad técnica, hay dos que destacan sobre el resto. La primera tiene que ver con la posibilidad de lograr un consenso sobre lo éticamente correcto; la segunda sobre si el uso de estas técnicas va en contra de la autonomía de las personas.

Y es que, a los problemas tradicionales que acompañan la discusión sobre la mejora humana se añade, en este asunto, el de cual es la moral que se toma como referencia. A las propuestas de mejora moral se las ha criticado por no

tener en cuenta la complejidad de la moral (Jotterand, 2022, 13 y ss.) o por estar basadas en un individualismo abstracto que no tiene en cuenta algo tan importante para la moral como es el contexto (Paulo/Bublitz, 2019, 95 y ss.).

En la actualidad, podríamos resolver estas cuestiones tomando como referencia a los derechos humanos. De alguna manera el discurso de los derechos puede ser contemplado como el referente de la mejora. Buena prueba de ello lo constituye el papel que estos representan en la educación. Como es sabido, el artículo 26 de la Declaración Universal de los Derechos Humanos, referido a la educación, en su punto 2 señala: "La educación tendrá por objeto el pleno desarrollo de la personalidad humana y el fortalecimiento del respeto a los derechos humanos y a las libertades fundamentales; favorecerá la comprensión, la tolerancia y la amistad entre todas las naciones y todos los grupos étnicos o religiosos, y promoverá el desarrollo de las actividades de las Naciones Unidas para el mantenimiento de la paz". Precisamente el ejemplo de la educación como herramienta de mejora moral es normalmente utilizado, como hemos visto, por los defensores de estas prácticas (Walker, 2009, 27 y ss.).

Pero claro, todos somos conscientes de que los derechos pueden ser interpretados de muchas maneras y, por tanto, pueden servir para justificar decisiones opuestas. Por eso la referencia a los derechos no es suficiente; es necesario optar por una teoría de los derechos.

Por otro lado, está la cuestión de la autonomía. Seguramente sea John Harris, y su defensa de la libertad de caer, quien ha manifestado de manera más conocida esta crítica. Sin tener esta libertad, dice Harris, no es posible actuar de forma inmoral ni discernir entre el bien y el mal: "La autonomía seguramente requiere no solo la posibilidad de caer sino la libertad de elegir caer, y esa misma autonomía nos da autosuficiencia; suficiente para haber permanecido libre para caer" (Harris, 2011, 103).

En cualquier caso, es posible, de manera general, diferenciar dos grandes tipos de herramientas, en el campo de la tecnología, que pretenden esta mejora moral: las basadas en la Biotecnología y las basadas en la Inteligencia Artificial.

Las herramientas biotecnológicas son las más tradicionales y pretenden fortalecer motivaciones morales o limitar motivaciones inmorales (Persson/Savulescu, 2012). No obstante, se trata de herramientas que han sido objeto de críticas al considerarse que erosionan la autonomía, reduciendo considerablemente las opciones de comportamiento y afectando a la autodeterminación, el autogobierno y la autenticidad (Lara Sánchez, 2021).

En el ámbito de la Inteligencia Artificial existen propuestas que utilizan la robótica con el objetivo apoyar y desatascar decisiones sin influir necesariamente

en su contenido final (Klincewicz, 2019, 425 y ss.), o influyendo desde el apoyo en teorías éticas, como la de Rawls (Borenstein/Arkin, 2016, 31 y ss.).

En todo caso, en los últimos años estamos asistiendo a la proliferación de herramientas, en muchos casos resultado de la combinación de la Inteligencia artificial y la Neurociencia (Farisco/Evers/Salles, 2022), que pretenden esa mejora moral y que buscan superar la crítica de la autonomía.

Entre ellas destacan aquellas que se basan en teorías de las decisiones como la de Thaler y Sunstein (2009) con su paternalismo libertario y los *nudges*. Como es sabido, estos dos autores se refieren a los pequeños empujones o impulsos, a través de los cuales instituciones y personas, como arquitectos de las decisiones (aquellos que tienen la responsabilidad de organizar el contexto en el que tomamos decisiones), estimulan, incentivan o encaminan la decisión sin traicionar la libertad de elección.

El paternalismo libertario se basa en "la convicción de que, en general, la personas deben ser libres para hacer lo que desean, y para desvincularse de los acuerdos desventajosos si lo prefieren"; y al mismo tiempo, en la consideración de que "es legítimo que los arquitectos de las decisiones traten de influir en la conducta de la gente para hacer su vida más larga, más sana y mejor". Un *nudge* es "cualquier aspecto de la arquitectura de las decisiones que modifica la conducta de las personas de una manera predecible sin prohibir ninguna opción ni cambiar de forma significativa sus incentivos económicos" (Thaler y Sunstein, 2009, 19 y 20).

Al paternalismo libertario se le ha criticado, entre otras cosas, la posible existencia de sesgos y prejuicios en su realización o la existencia de conflictos de intereses (Grüne-Yanoff, 2016, 463 y ss.), pero aun así, esta teoría se ha trasladado a la ética dando lugar a la ética de los impulsos o de la influencia (Sunstein, 2016), consistente en la realización de recomendaciones con el objetivo de influir en las personas a la hora de tomar decisiones éticas.

La teoría de los pequeños empujones está en la base de distintas propuestas para la mejora moral en el ámbito de las aplicaciones tecnológicas. Dentro de ellas, es posible diferenciar entre propuestas que poseen un carácter predominantemente sustantivo y aquellas que poseen un carácter predominantemente procedimental.

Entre las primeras, pueden citarse los modelos computacionales Truth-Teller y Sirocco (McLaren, 2006) y el programa MeEthEx, que es una especie de asesor ético en el campo de la medicina para ayudar a resolver dilemas éticos, y que se apoya en los principios de ética biomédica de Beauchamp y Childress, en la teoría del deber prima facie de Ross y en el equilibrio reflexivo de Rawls (Anderson/Anderson/Armen, 2005).

Entre las segundas puede citarse el modelo computacional diseñado por Robbins y Wallace (2007, 1571 y ss.), que se configura como una herramienta para ayuda en la toma de decisiones basado en el modelo creencia-deseo-intención y en la resolución colaborativa de problemas, simulando diferentes papeles (asesor, facilitador de grupo, entrenador de interacción y pronosticador). O, también la propuesta de F. Lara, primero en Lara/Deckers (2020) y luego en Lara (2021), de un asistente virtual para fortalecer nuestra moralidad potenciando la autonomía personal. Se trata de un asistente basado en el método dialéctico socrático si bien proyectado hacia el aprendizaje moral. El asistente, que denomina SocrAI, lo que pretende es formar al usuario no tanto en principios éticos sustantivos sino en pautas generales sobre cómo razonar mejor. Así SocrAI está presidido por tres grandes principios: la orientación educativa, la plena participación del usuario y la neutralidad de valores.

Podemos incluir en este ámbito la aplicación del Markkula Center for Applied Ethics de Santa Clara University (https://www.scu.edu/ethics/ethics-resources/a-framework-for-ethical-decision-making/). Para esta aplicación, la ética está compuesta por "normas y prácticas que nos dicen cómo deben actuar los seres humanos en las situaciones en las que se encuentran: como amigos, padres, hijos, ciudadanos, empresarios, profesionales, etc. La ética también se ocupa de nuestro carácter. Requiere conocimientos, habilidades y hábitos". Así, se subraya que la ética no es lo mismo que sentimientos, que religión, que cumplir el Derecho o normas culturales, y que tampoco es ciencia. A partir de ahí, se muestra una aplicación que ayuda a tomar decisiones éticas tomando como referencia lo que denomina como seis lentes éticos: de los derechos (basada en la dignidad igual de todos y en la existencia de derechos morales); de la justica (social, distributiva, correctiva, retributiva y restaurativa); de la utilidad (la acción ética es la que produce el mayor equilibrio entre el bien y el daño para la mayor cantidad posible de partes interesadas); del bien común (la vida en comunidad es un bien en sí mismo y nuestras acciones deben contribuir a esa vida); de la virtud (las acciones éticas deben ser consistentes con ciertas virtudes ideales que prevean el pleno desarrollo de nuestra humanidad); y de la ética del cuidado (tener en cuenta las relaciones, preocupaciones y sentimientos de todas las partes interesadas). El marco para la toma de decisiones éticas implica cinco pasos: a) identificar los problemas éticos; b) conocer los hechos; c) evaluar acciones alternativas; d) elegir una opción para la acción y probarla; e) implementar la decisión y reflexionar sobre el resultado**.**

En todo caso, la aplicación de la teoría de los empujones al ámbito tecnológico en combinación con el Big Data, ha dado lugar a los llamados *hiper nudges*, objeto de crítica por lo que conllevan de manipulación (Yeung, 2017, 118). Y es que los problemas de fondo de la mejora moral en relación con estas

propuestas siguen estando presentes (Agar, 2015, 343 y ss.), pero es cierto que las tecnologías convergentes y entre ellas, con mucha fuerza, la Inteligencia Artificial, están construyendo propuestas relevantes y singulares, que pretenden, como he señalado, superar la crítica del paternalismo, y que conviene tener en cuenta y discutir. Algunas de ellas, como ya adelanté, llegan desde reconocidos transhumanistas (Giubilini, Savulescu, 2018, 169 y ss.).

2. LA SUFICIENCIA DEL ENFOQUE DE DERECHOS HUMANOS

Un segundo reto que trataré es el de la suficiencia del enfoque de derechos humanos. Se trata de un reto que posee diferentes proyecciones, habiendo encontrado ya algunas de ellas una respuesta con cierto consenso en la comunidad internacional.

En efecto, dentro de lo que denomino como suficiencia del enfoque de derechos humanos es posible hacer referencia a tres cuestiones distintas pero relacionadas. La primera de ellas, es la de si los derechos constituyen los referentes adecuados para el tratamiento de las tecnologías convergentes. Se trata de una cuestión que parece haber sido resuelta ya en un sentido positivo.

En 2015 publicaba *Una mirada a la robótica desde los derechos humano* (De Asís, 2015) donde defendía ya que las respuestas a las tecnologías convergentes debían hacerse utilizando un enfoque de derechos humanos. Hoy la presencia de los derechos humanos a la hora de abordar estas cuestiones es un hecho. Lo era desde hace tiempo en el campo de la Genética y lo ha empezado a ser, sobre todo a partir de 2018, en los campos de la Inteligencia Artificial y la Robótica.

Michelle Bachelet, Alta Comisionada de las Naciones Unidas para los Derechos Humanos, en su discurso "Derechos humanos en la era digital. ¿Pueden marcar la diferencia?" de 17 de octubre de 2019, afirma: "Es esencial que en esta era digital prestemos especial atención a los derechos humanos… La revolución digital plantea un considerable problema de derechos humanos a escala mundial. Sus beneficios indudables no anulan sus riesgos evidentes". La Alta Comisionada se pregunta: "¿Y abordamos estos desafíos mediante la ética o mediante los derechos humanos?". Pues bien, Bachelet señala al respecto que "los códigos éticos y el cumplimiento voluntario no constituyen, por sí mismos, una respuesta suficientemente enérgica para la escala del problema que afrontamos… No hay ningún segmento de la revolución digital que no pueda y no deba examinarse desde una perspectiva de derechos humanos".

Ahora bien, como ya he señalado, tomar como referencia los derechos no parece suficiente, sino que es necesario optar por una teoría de los derechos. Los dos asuntos tradicionales de la teoría de los derechos, esto es, su concepto y fundamento, adquieren de nuevo relevancia.

A partir de ahí, cobran relevancia las otras dos cuestiones. Por un lado, la de si el actual marco de los derechos humanos es suficiente o es necesario reconocer nuevos derechos; por otro, la de si, además de los derechos, es necesario utilizar otras herramientas.

2.1. ¿Son necesarios nuevos derechos?

La reflexión ético-jurídica sobre las tecnologías convergentes han dado lugar a la demanda de nuevos derechos humanos. Algunos de ellos se ha propuesto en el marco de la genética (Morente, 2014, 195 y ss.), otros en el de la Inteligencia Artificial (Laukyte, 2021, 183 y ss.), otros en el de las neurotecnologías (Ienca/Andorno, 2017; Yuste/Genser/Herrmann, 2021, 154 y ss.).

Pues bien, parece necesario reflexionar sobre la necesidad del reconocimiento de nuevos derechos. Los defensores afirman que el discurso actual de los derechos no es suficiente para protegernos de las amenazas de las tecnologías convergentes. Sin embargo, está por ver si eso es así (Bublitz, 2022, 7) y si una reinterpretación de este discurso podría solucionar esta situación.

Los defensores de los neuroderechos probablemente contestarán que no, ya que las aplicaciones neurotecnológicas plantean problemas nuevos, antes insospechados (Ienca/Andorno, 2017). Sin embargo, si nos vamos a uno de los grandes instrumentos de protección de los derechos que es por cierto el instrumento jurídico que más tiempo lleva dialogando con la Neurociencia, el Derecho penal, tal vez esto no esté tan claro y la evolución actual de este Derecho caracterizada, entre otras cosas, por la desformalización (Muñoz/Conde, 2011, 7), permita integrar estos problemas en los bienes jurídicos presentes en los códigos.

También parece necesario tener en cuenta el problema tradicionalmente asociado a las propuestas de nuevos derechos y que se conoce como el problema de la inflación (cuantos más derechos reconocidos menos protección especial). Se trata en todo caso de una cuestión que tampoco está clara y que tiene que ver finalmente con el establecimiento de un buen sistema de garantías.

Igualmente es necesario analizar la consistencia de las propuestas. Y es que a las propuestas de los neuroderechos se les ha criticado precisamente su consistencia (Borbón/Borbón/Laverde, 2020, 152) y el "blanqueamiento" que hacen en relación con la mejora (Bublitz, 2022).

Por último, en el caso de que consideremos que se trata de propuestas que deben ser tenidas en cuenta, habrá que determinar un catálogo fundamentado y con un sistema de garantías que los haga eficaces.

En cualquier caso, un examen de las diferentes propuestas de nuevos derechos, tanto en el ámbito de los derechos digitales como en el de los neuroderechos, vuelve a poner de manifiesto cómo los problemas de la identidad y de la mejora siguen siendo los grandes temas a discutir. En efecto, de los cinco grandes grupos de neuroderechos a los que se refiere por ejemplo Ienca (Ienca, 2021), que coinciden en gran medida con los propuestos por Yuste y otros (Yuste/Genser/Herrmann, 2021), tres son fácilmente reconducibles a derechos y garantías ya existentes (libertad de pensamiento, privacidad e integridad). Sin embargo, el derecho a la identidad, que tiene una proyección *sui generis* en el campo de los derechos digitales a través de la identidad digital, no ha encontrado una plasmación jurídica en las Constituciones (salvo en la portuguesa). Y, el derecho al igual acceso a la mejora, sigue planteando la discusión previa sobre qué mejoras son éticamente aceptables.

2.2. ¿Son necesarias otras herramientas?

La discusión relativa a la necesidad o no de reconocer nuevos derechos, en el marco de la reflexión sobre la suficiencia del discurso de los derechos, ha ido acompañada de otra que trata sobre la oportunidad de utilizar otras herramientas, más allá de los derechos. En algunas de las propuestas sobre los nuevos derechos, podemos encontrar ejemplos de ello.

Por ejemplo, Yuste, Genser y Herrmann, para el reconocimiento y la satisfacción de los neuroderechos, proponen una serie de medidas, tres a corto plazo (destinadas a construir una definición consensuada de neuroderechos y con ello consolidar la investigación en neurotecnología y las prácticas regulatorias) y cuatro a largo plazo (destinadas a desarrollar tanto un marco para la protección y promoción de los neuroderechos como un mecanismo para monitorear las actividades de los países sobre neurotecnología). Las medidas a corto plazo son: a) la creación de una Comisión de Expertos en Derecho y Ciencia Internacional sobre Neuroderechos en Naciones Unidas; b) el nombramiento por Naciones Unidas de expertos altamente calificados para servir como asesores especiales sobre neuroderechos a organizaciones, instituciones e industria; c) el matenimiento de consultas periódicas con países clave por parte de los asesores y la Comisiónos. Las medidas a largo plazo son: a) la creación de un nuevo tratado o de un protocolo adicional a los tratados existentes para incorporar los neuroderechos; b) la elaboración de Comentarios generales sobre neuroderechos por parte de los Comités de seguimiento de los tratados; c) el nombramiento de un Relator especial sobre el impacto de la neurotecnología en los derechos humanos; d) la creación de una agencia especializada para coordinar las actividades globales de neuroderechos y ayudar a codificar los neuroderechos en un tratado internacional de derechos humanos.

Por su parte, la Oficina del Alto Comisionado de Naciones Unidas para los Derechos Humanos, en su informe sobre *El derecho a la privacidad en la era digital* (2021), ha señalado que los Estados deben establecer mecanismos de supervisión y reparación relacionados con la privacidad. La UNESCO, a través de su Comité Internacional de Bioética, emitió un informe el 15 de diciembre de 2021 sobre las Cuestiones Éticas de la Neurotecnología, en el que se encuentran también una serie de sugerencias generales y otras dirigidas a la propia UNESCO, a los Estados, a los investigadores, a las industrias, a los medios de comunicación e incluso al público en general. Entre las generales se señala: "a) Agregar protocolos a los tratados internacionales, como la Declaración Universal de los Derechos Humanos, para abordar los desafíos que plantean las neurotecnologías. b) Reforzar la Declaración Universal de los Derechos Humanos, considerando que la neurotecnología desafía los derechos humanos existentes y que se requerirán nuevas garantías en función de las posibilidades de vulneración. c) Elaborar una Nueva Declaración Universal de Derechos Humanos y Neurotecnología". Las recomendaciones a los Estados se resumen en la "concesión de un estatus positivo a los neuroderechos", velando por que "sus leyes fundamentales reconozcan y garanticen claramente la integridad física y psíquica que permita a las personas el pleno goce de su identidad personal, y el derecho a obrar de manera autónoma, y que sólo la ley pueda establecer los requisitos para limitarlo". Y en relación con el público se recomienda: "a) Enfatizar que cada individuo es el propietario de los datos que se recopilan de él o ella, y que solo pueden ser utilizados, publicados o comercializados en circunstancias excepcionales y solo con el consentimiento informado explícito. b) Tomar conciencia de los beneficios y riesgos potenciales de la neurotecnología, especialmente cuando afecta la integridad individual, influye en la percepción o induce a la toma de decisiones, y participa en debates públicos y otras acciones para examinar posibles abusos. c) Involucrarse en temas de neuroética y neuroderechos, de forma individual o mediante la formación de grupos de interés. d) Usar medios legales, incluyendo leyes y presión pública, para prevenir el abuso potencial de la neurotecnología por parte del gobierno, las agencias públicas o el sector privado".

Como podrá observarse, estas medidas cobran sentido al hilo del reconocimiento de nuevos derechos y vuelven a plantear la necesidad de que esta reflexión forme parte de la agenda política y científica.

3. UNAS REFLEXIONES FINALES

El tratamiento de las tecnologías convergentes requiere de debate, concienciación y formación entorno a su importancia. También requiere de la participación de la ciudadanía en la determinación de su sentido y alcance; y,

como no, de la continua atención a los destinatarios principales de los derechos: las personas en situación de vulnerabilidad.

En este sentido, es de celebrar la aparición de informes como el del Relator Especial sobre los derechos de las personas con discapacidad, Gerard Quinn, presentado en febrero de 2022 al Consejo de Derechos Humanos en cumplimiento de su resolución 44/10, titulado "La inteligencia artificial y los derechos de las personas con discapacidad". En este informe, el relator subraya que la inteligencia artificial puede ser una fuerza para el bien de las personas con discapacidad, especialmente si se vincula a la implementación de la CDPD, pero también puede tener efectos negativos importantes. Por eso, es necesario mantener un debate para examinar el equilibrio entre los riesgos y las oportunidades que presenta la inteligencia artificial en el contexto de la discapacidad. En ese debate, las personas con discapacidad y las organizaciones que las representan deben ser los protagonistas. Además, como es habitual, en el informe se especifican recomendaciones a los Estados, a las instituciones nacionales de derechos humanos, a las empresas y el sector privado, al sistema de Naciones Unidas y al Comité sobre los Derechos de las Personas con Discapacidad.

Que los derechos humanos presidan la reflexión sobre la incidencia de las tecnologías en nuestras sociedades no implica necesariamente, como algunos creen, la imposición de límites al desarrollo de éstas. Al mismo tiempo, atender a las situaciones de vulnerabilidad no supone tampoco manejar una visión optimista de las tecnologías porque, como ya hemos apuntado, en el tratamiento de la vulnerabilidad, éstas, en ocasiones, pueden pretender acabar con la diversidad.

Recientemente, Edward Ashford Lee, profesor de la Universidad de California en Berkeley, se ha referido al problema de la regulación de las tecnologías convergentes preguntándose si estamos perdiendo su control (Lee, 2022). Y esta pregunta la contesta el informático puertorriqueño afirmando que no, pero porque nunca lo hemos tenido y no se puede dejar de tener aquello que nunca se ha tenido. Lee justifica su posición diferenciando entre dos formas de contemplar la tecnología que denomina como creacionismo digital y coevolución. Para la primera, que considera que es el enfoque estándar, la tecnología es el resultado de un diseño inteligente de arriba hacia abajo; cada tecnología es el resultado de un proceso basado en una decisión humana. Sin embargo, para la segunda, que es para Lee la manera correcta de contemplar la tecnología, ésta se desarrolla de una forma horizontal, siendo los tecnólogos algo así como unos agentes de mutación, estando el éxito y el desarrollo de sus productos determinados por herramientas y entorno. Para Lee, lo importante es centrarse en el uso, y no tanto en el diseño. Y así, considera que podría ser tan

efectivo aprobar normas que se centren en educar a la sociedad que aprobar normas que regulen a quienes producen tecnología.

Traigo aquí la reflexión de Lee porque a lo largo de este breve trabajo he señalado cómo los derechos humanos están presentes en la reflexión sobre la incidencia social de las tecnologías convergentes y, también como, al hilo de la existencia de propuestas de nuevos derechos, es necesario promover un debate amplio e interdisciplinar. Sin embargo, seguimos sin prestar una excesiva atención a esa otra herramienta que apunta Lee y que seguramente sea mucho más eficaz: la educación y la formación en derechos humanos.

Y es que, en este punto, resulta paradójico observar como se promueve una formación cada vez mayor en el campo tecnológico y, al mismo tiempo, se discute la oportunidad de formar en los derechos humanos. Asistimos a una proliferación de cursos dirigidos a toda la población sobre tecnología y en pocos de ellos se abordan cuestiones éticas. Y todo ello se justifica señalando que la cuestión de los derechos es ideología, presumiendo que en cambio la tecnología es neutra.

Sin embargo, como se han encargado de demostrar de nuevo Sunstein y Thaler, el diseño neutral no existe (Sunstein/Thaler, 2009, 17) y, efectivamente, los derechos humanos son una toma postura, la de la dignidad humana. Una toma de postura que debe estar abierta al progreso y a la diversidad (De Asís/Laukyte, 2020), y que no está reñida necesariamente con algunas concepciones transhumanistas (Liedo/Rueda, 2021, 215 y ss.), ni completamente cerrada a planteamientos de mejora, pero que debe ser fiel a su principal objetivo: el libre desarrollo de la personalidad de todas las personas.

4. BIBLIOGRAFÍA

Agar, Nicholas (2015), "Moral Bioenhancement is Dangerous", en: *Journal of Medical Ethics* 41, 343-345.

Agustín de Hipona (2002), "El espíritu y el alma", en: *Obras completas de San Agustín. XLI: Escritos atribuidos*, BAC, Madrid.

Anderson, Michael, Susan Anderson, Chris Armen (2005), "MedEthEx: Toward a Medical Ethics Advisor", en: *AAAI Fall Symposium: Caring Machine*, Disponible en https://www.aaai.org/Papers/Symposia/Fall/2005/FS-05-02/FS05-02-002.pdf. (18/4/2022).

Barranco Avilés, Mª del Carmen (2016), *Condición humana y derechos humanos*, Dykinson, Madrid.

Boire, Richard Glen (1999), "On cognitive liberty I", en: *Journal of Cognitive Liberties*, 1, 7-13.

Borbón Rodríguez, Diego, Diego Luisa Borbón, Jennifer Laverde (2020), "Análisis crítico de los NeuroDerechos Humanos al libre albedrío y al acceso equitativo a tecnologías de mejora", en: *Ius et Sciencia* 6, 2, 135-161.

Borenstein, Jason, Ron Arkin (2016), "Robotic Nudges: The Ethics of Engineering a More Socially Just Human Being", en: *Science and Engineering Ethics* 22, 31-46.

Bostrom, Nick (2005), "The Fable of the Dragon-Tyrant", en: *Journal of Medical Ethics* 31, No. 5, 273-277.

Bublitz, Jan Christoph (2022), "Novel Neurorights: From Nonsense to Substance", en: *Neuroethics*, 15, 1-15.

De Asís, Rafael (2001), *Sobre el concepto y el fundamento de los derechos: una aproximación dualista*, Dykinson, Madrid.

— (2013), *Sobre discapacidad y derechos*, Dykinson, Madrid.

— (2015), *Una mirada a la robótica desde los derechos humanos*, Dykinson, Madrid.

— (2022), "Sobre la propuesta de los neuroderechos", en: *Derechos y Libertades*, 22, 51-70.

De Asís, Rafael, Migle Laukyte (2020), "Transhumanismo y envejecimiento", en: *Soluciones tecnológicas para los problemas ligados al envejecimiento: cuestiones éticas y jurídicas*, 93-114.

Descartes, René (1989), "Meditaciones metafísicas", en: *Discurso del método. Meditaciones metafísicas*, Espasa-Calpe, Madrid.

Douglas, Thomas (2008), "Moral enhancement", en: *Journal of Applied Philosophy* 25, 3, 162-177.

Farisco, Michele, Kathinka Evers, Arleen Salles (2022), "On the Contribution of Neuroethics to the Ethics and Regulation of Artificial Intelligence", en: *Neuroethics* 15, 4, 1-12.

Ferrater Mora, José (1965a) "Persona", *Diccionario de Filosofía, T. II*, Editorial Sudamericana, Buenos Aires.

— (1965b) "Identidad", *Diccionario de Filosofía, T. I*, Editorial Sudamericana, Buenos Aires.

Giubilini, Alberto, Julian Savulescu (2018). "The Artificial Moral Advisor. The "Ideal Observer" Meets Artificial Intelligence", en: *Philosophy & technology*, 31, 2, 169-188.

Grüne-Yanoff, Till (2016), "Why behavioural policy needs mechanistic evidence", en: *Economics and Philosophy*, 32, 3, 463-483.

Hall, Melinda C. (2020), "Second Thoughts on Enhancement and Disability", en: Cureton, Adam, David T. Wasserman (eds.), *The Oxford Handbook of Philosophy and Disability*, Oxford University Press, Oxford.

Harris, John (2011), "Moral Enhancement and Freedom", en: *Bioethics* 25/2, 102-111.

— (2016), *How to be Good. The Possibility of Moral Enhancement*, Oxford University Press, Oxford.

Hopster, Jeroen (2022), "The Ethics of Disruptive Technologies: Towards a General Framework", en: De Paz Santana, J.F. *et al.* (eds), *New Trends in Disruptive Technologies, Tech Ethics and Artificial Intelligence. DiTTEt 2021. Advances in Intelligent Systems and Computing,* vol 1410, Springer, Cham, 1-12.

Ienca, Marcello, Roberto Andorno (2017), "Towards new human rights in the age of neuroscience and neurotechnology", en: *Life Sciences, Society and Policy*, 13, 5, 1-27.

Ienca, Marcello (2021), "On neurorights", en: *Frontiers in Human Neurosciencie*, 24 September, 1-11.

Jotterand, Fabrice (2022), *The Unfit Brain and the Limits of Moral Bioenhancement*, Palgrave Macmillan, Singapore.

Klincewicz, Michal (2019), "Robotic nudges for moral improvement through Stoic practice", en: *Techné: Research in Philosophy and Technology*, 23, 3, 425-455.

Lara Sánchez, Francisco (2016), "El imperativo ético de la mejora moral", en: *Gazeta de antropología*, núm. 32, 2. http://hdl.handle.net/10481/43306

— (2021), "Why a Virtual Assistant for Moral Enhancement When We Could have a Socrates?", en: *Science and engineering ethics*, 27, 4, 42, 1-27.

Lara, Francisco, Jan Deckers (2020), "Artificial Intelligence as a Socratic Assistant for Moral Enhancement", en: *Neuroethics* 13, 275-287.

Laukyte, Migle (2021), "Dignidad humana y nuevos derechos: el derecho a la Inteligencia Artificial", en: Llano Alonso, Fernando H., Joaquín Garrido (eds.), *Inteligencia Artificial y Derecho. El jurista ante los retos de la era digital*, Aranzadi, Pamplona, 183-199.

Lee, Edward A. (2022), "Are We Losing Control?", en: Werthner, Hannes *et al.* (eds.), *Perspectives on Digital Humanism*, Springer, Cham, 3-8.

Lema, Carlos (2015), "Intervenciones bioéticas de mejora, mejoras objetivas y mejoras discriminatorias: ¿de la eugenesia al darwinismo social?", en: *Anales de la Cátedra Francisco Suárez*, 49, 367-393.

Liedo Fernández, Belén, Jon Rueda (2021), "In Defense of Posthuman Vulnerability", en: *Scientia et Fides*, 9, 1, 215-239.

Llano Alonso, Fernando H. (2018), *Homo Excelsior. Los Límites Ético Jurídicos del Transhumanismo*, Tirant Lo Blanch, Valencia.

McLaren, Bruce M. (2006), "Lessons in Machine Ethics from the Perspective of Two Computational Models of Ethical Reasoning". Disponible en: https://www.researchgate.net/publication/228706785_Lessons_in_machine_ethics_from_the_perspective_of_two_computational_models_of_ethical_reasoning/figures?lo=1 Última consulta: 18 de abril de 2022.

Morente Parra, Vanesa (2014), *Nuevos retos biotecnológicos para los derechos fundamentales*, Comares, Granada.

Mosterín, Jesús (2013), "Naturaleza humana, Biología y Convención", en: *Estudios Públicos,* 131, 2013.

Muñoz Conde, Francisco (2011), "Protección de los derechos fundamentales en el Código penal", en: *Derecho y Cambio Social*, 22, 1-11.

Nussbaum, Martha (2002), "Genética y Justicia: Tratar la enfermedad, respetar la diferencia", en: *Isegoria*, 27, 5-17.

Paulo, Norberto, Jan Christoph Bublitz (2019), "How (not) to Argue for Moral Enhancement: Reflections on a Decade of Debate", en: *Topoi* 38, 95-109.

Peces-Barba, Gregorio (1989), "Sobre el fundamento de los derechos humanos: un problema de moral y derecho", en: Muguerza, Javier *et al., El fundamento de los derechos humanos*, Debate, Madrid, 265-277.

Persson, Ingmar y Savulescu, Julian (2008), "The Perils of Cognitive Enhancement and the Urgent Imperative to Enhance the Moral Character of Humanity", en: *Journal of Applied Philosophy*, 25, 3, 162-177.

— (2012), *Unfit for the future*. Oxford University Press, Oxford.

Piñar Mañas, José Luis (2018), "Identidad y personas en la sociedad digital", en: AA.VV., *Sociedad digital y Derecho*, BOE, Madrid, 95-112.

Robbins, Russel W., William Wallace (2007), "Decision support for ethical problem solving: A multi-agent approach", en: *Decision Support Systems*, 43, 1571-1587.

Sandel, Michael (2007), *Contra la perfección*, Marbot, Barcelona.

Savulescu, Julian, Hannah Maslen (2015), "Moral Enhancement and Artificial Intelligence: Moral AI?", en: Romportl, Jan, Eva Zackova, Josef Kelemen (eds.), *Beyond Artificial Intelligence*, Springer International Publishing, Cham, 79-95.

Sententia, Wrye (2004), "Neuroethical considerations: cognitive liberty and converging technologies for improving human cognition", en: *Annals of the New York Academy of Science*, 1013, 221-228.

Shoemaker, Sydney (1996), *The first-person perspective and other essays*, Cambridge University Press, Cambridge.

Sunstein, Cass R. (2016), *The Ethics of Influence*, Cambridge University Press, 2016.

Thaler, Richar, Cass R. Sunstein, C. (2009), *Nudge: Improving decisions about health, wealth, and happiness*, Penguin, London.

— (2009), *Un pequeño empujón*, Taurus, 2009.

Walker, Marl (2009), "Enhancing genetic virtue: A project for twenty-first century humanity?", en: *Politics and the Life Sciences*, 28, 2, 27-47.

Wiggins, David (1980), *Sameness and Substance*, Harvard University Press, Cambridge.

Yeung, Karen (2017). "Hypernudge: Big Data as a mode of regulation by design", en: *Information, Communication & Society*, 20, 1, 118-136

Yuste, Rafael, Jared Genser, Stephanie Herrmann (2021), "It's Time for Neuro-Rights", en: *Horizons, Center for International Relations and Sustainable Development*, 18, 154-164.

CAPÍTULO II

LA PROBLEMÁTICA DE LOS SESGOS ALGORÍTMICOS (CON ESPECIAL REFERENCIA A LOS DE GÉNERO). ¿HACIA UN DERECHO A LA PROTECCIÓN CONTRA LOS SESGOS?

NURIA BELLOSO MARTÍN
Universidad de Burgos
nubello@ubu.es

1. INTRODUCCIÓN

Numerosos estudios ya han advertido de los riesgos y posibles vulneraciones de los que algunos derechos fundamentales son objeto desde la irrupción del nuevo espacio digital. Las discriminaciones y sesgos que muchas veces presiden las decisiones y acciones humanas -a veces de forman inconsciente-, se proyectan también en la red a través del software y de sistemas de Inteligencia Artificial (en adelante, IA) (Lettieri, 2020; Vantin, 2021; Menéndez Sebastián, 2021). Aunque, de entrada, podría pensarse que el espacio digital conlleva la ventaja de ser un espacio neutro y aportar objetividad, de manera que promueva relaciones igualitarias y equitativas, alejado de la discriminación que, en ocasiones, preside juicios y elecciones humanas, sin embargo, no es así, habiendo dado lugar a amenazas que adoptan nuevas formas, como son los sesgos algorítmicos. La lucha contra la discriminación ha sido una de las coordenadas que han presidido los derechos humanos (Bellver Capella, 2021) y, si tales discriminaciones se producen en el ámbito tecnológico, hay que establecer las regulaciones oportunas (legales, tecnológicas y éticas) para evitarlas (Kleinberg, 2018). Basta recordar que los ejes sobre los que se erige la Carta ética europea, de 2018, en la que, junto a la calidad y seguridad, transparencia, y control del usuario, se establece también la no-discriminación, la imparcialidad y la equidad.

No hay consenso sobre el concepto de los sesgos ni tampoco hay acuerdo sobre si son negativos en su totalidad o debe aceptarse un mínimo de sesgos, precisamente para hacer posible las diferencias y, con ello, la equidad y justicia, tratando de forma diferente a quien sea diferente. Hay sesgos que se heredan de sesgos históricos, otros, se han ido construyendo socialmente. Son sesgos escurridizos, resbaladizos, sutiles en muchas ocasiones, cuyas consecuencias son extremadamente negativas para las personas a quienes afectan -mujeres en el caso de discriminación de sexo, personas mayores en el caso del edadismo, racismo en el caso de personas de otra raza distinta a la blanca, estrato social, religión, y otros-.

Las "injusticias" algorítmicas (discriminación, sesgos, perfilado) llevan a (re)pensar una especie de nuevo contrato social, ahora "tecno-social". Sin embargo, la mayor parte de regulación al respecto es de *soft law*, es decir, se reduce en su mayoría a textos, Informes y Cartas (éticas) -como los emanados de la UNESCO, de instituciones de Latinoamérica y de la Unión Europea para regular diversos aspectos "éticos" de la IA-. A modo de ejemplo, la Declaración de Principios Éticos para la IA de Latinoamérica IA-LATAM para el diseño, desarrollo y uso de la Inteligencia Artificial, en su punto octavo, hace referencia expresa a los sesgos: **"**Evitar los sesgos e impactos injustos en las personas, en particular las relacionadas con características sensibles como la raza, el origen étnico, el género, la nacionalidad, los ingresos, la orientación sexual, la capacidad y las creencias políticas o religiosas".

Dado que la IA trabaja a través de la padronización de acciones y del aprendizaje automático a partir de los datos, hay que controlar esas fases en las que surgen sesgos que después se replicarán y que podrán acabar arrojando un resultado discriminatorio para algún individuo o colectivo (por razón de raza, religión, etnia, sexo, clase social) (Myers West/Whittaker/Crawford, 2009). El reto -tanto tecnológico como jurídico- es tanto el de identificar los sesgos como el de diseñar los procedimientos necesarios para poder minimizarlos o neutralizarlos.

El presente estudio tiene por objeto analizar los sesgos que se producen al aplicar una IA, y que tienen como consecuencia una discriminación y una vulneración de la igualdad. La IA puede ayudar a identificar y reducir estos sesgos, pero también puede aumentar el problema de la inclusión y la diversidad si no están integradas en todo el ciclo de vida del sistema de la IA. En primer lugar, examinaré las actuaciones y factores que provocan los sesgos algorítmicos en general para centrarme después en los sesgos de género en particular. Ya advierto de entrada que, a mi juicio, no se trata de un error o un mal funcionamiento del programa, sino de una fase o un conjunto de fases previas o posteriores que no se han implementado, o que no se han desarrollado y ejecutado adecuadamente. En segundo lugar, examinaré la tipología de sesgos algorítmicos y su incidencia en una condición concreta, como es el género (Bernheim/Vincent, 2019). En tercer lugar, tras revisar el estado de la cuestión en el ámbito internacional, se formularán algunas propuestas de uso de la aplicación de la IA con enfoque de género, y ello, no sólo para minimizar o eliminar los sesgos sino para emprender un uso proactivo de la IA para lograr la equidad de género, aprovechando la potencialidad que tiene la IA para el logro de tal empeño. Terminaré con unas reflexiones finales y unas propuestas.

2. ¿POR QUÉ LO ATRIBUYEN A UN "ERROR" DE LA INTELIGENCIA ARTIFICIAL CUANDO SE TRATA DE SESGOS (ALGORÍTMICOS)?

El tratamiento de los sesgos algorítmicos -bien se apliquen inconscientemente, bien se oculten a propósito- constituye un área activa de investigación en los avances de Inteligencia artificial (IA) (Herranz, 2019).

Comienzo subrayando la falta de una definición "jurídica" sobre qué sea sesgos. En este sentido, este es uno de los aspectos que se ha señalado por los expertos que debería de contemplar la versión definitiva de lo que, ahora, es la Propuesta de Reglamento de la Comisión Europea del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de IA (Ley de Inteligencia artificial) de 2021. El proyecto de Reglamento, en su punto 3.5, sobre el uso de una IA fiable en cuanto a los derechos fundamentales, hace mención expresa a la Carta de derechos fundamentales de la UE, y a la promoción de la protección de derechos como el de no discriminación (art.21) y a "minimizar el riesgo de adoptar decisiones asistidas por IA erróneas o sesgadas". Mientras tal Propuesta se consolida en Reglamento, hay que partir ahora de clarificar a qué se hace referencia con el término "sesgo" (*bias*) para después analizar de dónde derivan y qué tipos de sesgos algorítmicos existen. Un sesgo hace referencia a una inclinación o desviación hacia algo. H. Ramírez define los sesgos como:

> Son aquellas creencias inconscientes que todos tenemos sobre hombres y mujeres, basadas en los estereotipos socioculturales con los que nos hemos educado e interiorizado. Se trata, por tanto, de ideas, predilecciones o prejuicios inconscientes, que se activan la mayoría de las veces de forma automática, porque nuestro cerebro funciona a través de la minimización del esfuerzo cognitivo y los estereotipos permiten tomar decisiones de forma más rápida (Ramírez, b, 2021).

Esta definición se puede completar con la que ofrece la Real Academia Española, relativa al área de la estadística, que lo conceptualiza como: "Error sistemático en el que se puede incurrir cuando al hacer muestreos o ensayos se seleccionan o favorecen unas respuestas frente a otras". En esta acepción, se configura con una concepción negativa, que es la que generalmente se tiene con respecto a los sesgos y, de ahí, las investigaciones que giran sobre los procedimientos para neutralizarlos o eliminarlos.

Se hace necesario indagar porqué los sesgos -que se encuentran en valores, creencias, normas, culturas-, también se hallan en los algoritmos con los que trabaja la Inteligencia Artificial. ¿Obedecen a que están incorporados a los datos?, ¿son consecuencia de un defectuoso entrenamiento de los algoritmos? ¿existen porque el programador los ha incluido? Tanto los conjuntos de datos

como las decisiones computarizadas ofrecen una representación imperfecta del mundo, ya que constituyen juicios humanos que reflejan una visión sobre cómo es el mundo. Los sesgos algorítmicos no ocurren de manera espontánea. Los algoritmos de IA basados en datos, no producen sesgos pero sí pueden reproducirlos sin la adecuada intervención humana y ello en tres de las fases principales: en la recolección de los datos, porque tales datos recopilados reflejen prejuicios ya existentes; en la preparación de datos de entrenamiento (a la hora de seleccionar y procesar los atributos que le proporcionamos al algoritmo); y en la toma de decisiones (las propuestas y decisiones que se adoptan a lo largo de todo el ciclo de vida del desarrollo inteligente).

Los sistemas de Inteligencia Artificial se basan en correlaciones, en diseño de perfiles (Hao, 2019), en búsqueda de patrones, y precisamente esa es una de las grandes diferencias con la inteligencia humana -y lo que, a la postre, puede derivar en una padronización de modelos erróneos que acabe desembocando en los sesgos -como es el caso del ya conocido sesgo por raza del modelo COMPAS -puesto de manifiesto por un estudio de Propublica- en el ámbito penal norteamericano, en la sentencia de 13 de julio de 216, Wisconsin, Suprema Corte, *State v. Loomis*, en la que se recurrió a la IA para definir la probabilidad de reincidencia de un acusado- aunque debe advertirse que algunos autores consideran que los algoritmos bien diseñados deberían poder evitar sesgos cognitivos de muchos tipos (Sunstein, 2018). Las personas diferenciamos correlaciones de causa-efecto, pero las máquinas no.

La problemática de los sesgos algorítmicos -tanto en su sentido técnico como jurídico- puede englobarse en el "Estado algorítmico de Derecho", como apunta Barrio Andrés (2020):

> Se sabe que los sistemas algorítmicos plantean diversas cuestiones relacionadas con la parcialidad, la injusticia y la discriminación en las decisiones que adoptan, así como con la transparencia, la explicabilidad y la rendición de cuentas en lo que respecta a su funcionamiento o la protección de los datos, la privacidad y otras cuestiones de derechos fundamentales, entre otras. Incluso podemos hablar de "absurdos algorítmicos" para calificar sus cálculos incorrectos.

La realidad es que los sesgos algorítmicos están presentes de distintas formas y en diversos momentos. Quienes no están versados en el tema, suelen culpar al propio sistema, como si hubiera ejecutado erróneamente el programa. No se trata de que la IA cometa errores, sino de que, o bien el diseño del programa no era el correcto; o bien la selección de los datos recopilados para entrenar al algoritmo haya sido incompleta; o incluso, porque la interpretación de los resultados ha sido equivocada. A este respecto, Merino apunta que parecería

lógico pensar que las IA no padecen algunos males típicamente humanos, como los prejuicios; y que no discriminan por la raza / nacionalidad / cultura. Pero la realidad es un poco más compleja. El problema no son los prejuicios, sino los datos, como una mala selección de los datos usados para entrenar a la IA. Por ejemplo, ni los programadores de Beauty.AI, ni los de Microsoft, IBM o Megvii, diseñaron sus algoritmos con la intención de discriminar a ningún grupo humano, sino que el problema de su proyecto residió en que, al haberse entrenado mayoritariamente con fotos de personas de raza blanca, la IA no vinculaba la piel oscura al ideal de "belleza humana" (Merino, 2019; Doshi, 2018).

Como resultado, la respuesta que facilite el algoritmo puede generar decisiones injustas, comportando discriminación o estigmatización hacia ciertos grupos de personas o colectivos. En cualquier caso, no se puede atribuir a la IA sin más que "se equivoca" y que, por ello, produce los denominados "sesgos algorítmicos".

2.1. Algunos casos de discriminación algorítmica por parte del sector público

En otros casos, la discriminación es buscada porque es lo que pretende el programa, como el sistema del "*social credits*", un proyecto que ha lanzado China con el fin de delinear la reputación del "buen ciudadano" y cuyo objetivo es discriminar por clase (Fico, 2018). Aquí también se puede traer a colación el caso del bono social de electricidad español, que es un descuento en la factura eléctrica para los "consumidores vulnerables y con menor capacidad económica para afrontar los precios elevados". La aplicación del bono social eléctrico dio lugar a numerosas quejas y reclamaciones porque el programa BOSCO que se utilizaba para la concesión, negaba la ayuda a personas que tenían derecho a la misma. La norma establece tres vías de entrada para acceder al descuento: por rentas bajas, familias numerosas y beneficiarios de una pensión mínima de incapacidad o jubilación y que no cuenten con otros ingresos. Esta última vía deja fuera a otro tipo de pensionistas, como las viudas, y a los pensionistas que reciban algún ingreso. En el caso de una viuda, la respuesta del programa BOSCO contiene dos afirmaciones: "no reúne los requisitos" e "imposibilidad de comprobar niveles de renta". Las dos eran falsas: sí los reúne y sí se han comprobado los niveles de renta. La plataforma Civio, en septiembre de 2019, interpuso recurso contencioso-administrativo -tal recurso se ha resuelto con una sentencia desestimatoria para Civio y con una condena en costas, y que ha sido apelada- ante la negativa del Consejo de Transparencia de obligar a hacer público el código del programa que decide quién resulta beneficiario del bono social eléctrico (Civio, 2019). Puede observarse que es frecuente la interrelación entre sesgo discriminatorio y falta de transparencia de cómo funciona el sistema -la dificultad de acceder al código fuente es también conocida en cuanto la legislación protege las patentes y propiedad intelectual-.

Otro caso de discriminación de ingresos se puede citar con respecto al programa SyRI que se utilizaba por el Gobierno de los Países Bajos. El denominado Sistema de Indicación de Riesgos (*Systeem Risico Indicatie,* SyRI) es un instrumento legal que el gobierno neerlandés utilizaba para prevenir y combatir el fraude. El sistema se basaba en la asignación del nivel de riesgo de que una determinada persona cometa fraude a los ingresos públicos, en función de una serie de parámetros analizados y relacionados entre sí. El Sistema se había configurado a partir de la denominada Ley de Organización de Implementación y Estructura de Ingresos (*Wet structuur uitvoeringsorganisatie en inkomen,* SUWI), cuyo artículo 65.2 permite la elaboración de informes de riesgos para evaluar el riesgo de que una persona física o jurídica haga un uso ilegal de fondos gubernamentales en el campo de la seguridad social y los esquemas relacionados con los ingresos públicos. SyRI se utilizaba en especial en barrios en los que el Ejecutivo considera problemáticos. Para calcular posibles irregularidades, los algoritmos enlazan todos los datos personales de sus residentes almacenados por instancias gubernamentales. Esos datos se comparan luego con el perfil de riesgo creado a partir de la información de otros ciudadanos que sí han delinquido. Observadas las similitudes, se confecciona una lista de nombres que las autoridades pueden conservar hasta dos años.

El Tribunal de Distrito de La Haya (Holanda) (de primera instancia) (*Rechtbank Den Haag*), en su Sentencia del 5 de febrero de 2020, estableció que el sistema algorítmico SyRI, que venía utilizando ese programa para evaluar el riesgo de fraude a la Seguridad Social o al Ministerio de Hacienda era contrario a la ley. Al dar como resultado un listado de nombres de potenciales sospechosos como "probables defraudadores", el modelo de riesgo elaborado en esos momentos por SyRI desembocaba en una estigmatización y discriminación de la ciudadanía (Fernández, 2020). Su implementación no presentaba garantías suficientes en cuanto a la proporcionalidad debida para evitar injerencia en la privacidad, tal y como establece el artículo 8 de Convenio Europeo de Derechos Humanos. El caso refleja dos caras que se enfrentan constantemente en el debate de si es lícito depositar nuestra confianza en algoritmos que, por su diseño o entrenamiento, pueden presentar sesgos ocultos.

Estos casos ponen de manifiesto que la aplicación de la IA a la Administración pública, con la consiguiente automatización -autónoma o supervisada por un humano- no ha eliminado el sesgo, sino que ha cambiado la posibilidad de sesgo humano por la incorporación del sesgo sistémico de los algoritmos (Menéndez Sebastián, 2021).

También en el ámbito privado se han utilizado aplicaciones de IA con sesgos. Basta recordar, en la Sentencia de 31 de diciembre de 2020, del Tribunal Ordinario de Bolonia (Italia) sobre el caso *Deliveroo*. La aplicación de la IA "Frank" realizaba la asignación de los encargos de pedidos gastronómicos. Se presentó una demanda por discriminación en los índices de fiabilidad y disponibilidad utilizados por el algoritmo (no tenía en cuenta razones de no disponibilidad) por lo que se penalizaba a los trabajadores en esa situación. Se dictaminó que había discriminación indirecta", inconsciencia y ceguera deliberada" de la empresa.

El problema ético que se produce en estos casos se entrecruza con un problema técnico. En los casos que afectan a la Administración pública, se debe ser especialmente vigilante con la salvaguarda de los derechos de los ciudadanos, de manera que los programas de IA que se elaboren y apliquen sean comprensibles (*explanaible AI*) y consigan escapar a la opacidad de la autoreferencia; y que refuercen la transparencia y la explicabilidad (*explicability*); por último, la supervisión humana final permitirá ejercer el último control.

2.2. Las fases de formación de un sesgo algorítmico

La discriminación puede aparecer en cualquiera de las fases: *pre-processing*, *in-processing*, y *post-processing*. Rementería diferencia cuatro fases para formar un algoritmo: la fase de diseño y entrenamiento; la fase de uso y aplicación; la de validación y, por último, la de presentación (Rementería, 2021; Soriano Arnanz, 2021). Comenzando por la primera fase, la de entrenamiento, en la mayor parte de las ocasiones, tales sesgos no obedecen a un propósito definido por el programador o por quien ha desarrollado el sistema sino precisamente a un defectuoso entrenamiento del aprendizaje de los algoritmos, ya que se entrenan con modelos estándar. Por ejemplo, si va a ser un algoritmo de imagen, el sistema deberá aprender a despixelar, poco a poco, entre millones de rostros, que suele ser el de "varón blanco occidental", lo que dificulta, a los motores de búsqueda de un programa, el reconocimiento y clasificación de imágenes de "varones de raza negra"; si se trata de un algoritmo de voz, el algoritmo deberá entrenarse con todo tipo de voces (graves, agudas) para que pueda reconocerlas y asistir verbalmente a lo que se le solicite. Una de las causas de los problemas que derivan de los métodos actuales de análisis de datos, particularmente antes del desarrollo del modelo, es que son costosos y no están estandarizados. Por ello, para desarrollar con más fiabilidad la IA, hay propuestas para evaluar conjuntos de datos basados en medidas de calidad estándar, tanto cualitativas como cuantitativas. Por ejemplo, el MIT ha puesto en marcha el *Dataset Nutrition Label Project*, que actualmente está trabajando en

agrupar dichas medidas en una Etiqueta de Alimentación de Sets de Datos que resulte fácil de usar (Holland *et al.*, 2018).

Una segunda fase es la de uso y aplicación. Los algoritmos trabajan principalmente buscando correlaciones que maximicen la capacidad de predicción, lo cual puede provocar, en ocasiones, resultados basados en relaciones espurias y acabar desembocando en conclusiones sesgadas. La importancia de estos sesgos dependerá, lógicamente, del contexto en el que se produzcan. A modo de ejemplo, no es lo mismo incurrir en sesgos en la traducción de un texto que en la concesión de un crédito. En esta fase se pueden presentar sesgos de distinto tipo: estadísticos, tecnológicos, errores de medida, culturales, estereotipos de lenguaje, cognitivos (los que identifican nuestros gustos), conscientes/ inconscientes, implícitos/ explícitos, deseables/ no deseables, de exclusión (al coger la muestra, se ha dejado fuera a una parte importante de la población).

Como explica Fernández, la posible existencia de sesgos no intencionados puede tener su origen en los datos, en el modo en el que se entrena el algoritmo o en la fase de creación o de ajustes por parte del programador. Es decir, por una parte, los algoritmos necesitan entrenarse con una gran cantidad de datos y, además, deben ser de calidad, esto es, representativos de toda la población. De lo contrario, se pueden producir situaciones en las que el sesgo de la muestra de entrenamiento se incorpora como un criterio que se ha de cumplir, lo que dificulta que se avance en la igualdad de oportunidades (por ejemplo, casos como la selección de *curricula* para puestos de trabajo o la concesión de créditos) (Fernández, 2019, 5); o también, en motores de búsqueda de imagen, si el algoritmo ha entrenado con rostros de hombres, apenas conseguirá identificar imágenes de mujeres).

Por otra parte, pueden producirse sesgos no deseados por el modo en que se diseña o funciona el algoritmo (por ejemplo, en el etiquetado de los datos de entrenamiento o en el modo en que evoluciona el algoritmo a medida que incorpora nueva información). Más adelante se aludirá al caso del *chatbot* experimental que lanzó una empresa tecnológica en una red social a fin de que aprendiera conversando con los usuarios, y que tuvo que retirarse al poco tiempo debido al tono inapropiado que habían adquirido sus mensajes según lo que había aprendido en pocas horas de intercambio de información con usuarios.

Una tercera fase es la de validación, en la que el programador hace de observador para comprobar el funcionamiento del algoritmo; la cuarta y última fase es la de presentación cuando, una vez finalizado el algoritmo, el programador lo presenta a los promotores del producto. Aquí también se pueden introducir sesgos de diverso tipo, como los financieros (si los

promotores consideran que es muy caro, por ejemplo, solicitarán que se abaraten los costes, lo que puede derivar en la introducción de algún sesgo), culturales, aptos para unos fines de marketing o publicidad, o de otro tipo. Resulta necesaria una auditoría final del algoritmo para controlar esos sesgos y verificar su buen funcionamiento -tanto jurídico como ético-.

Una máquina no comete errores, sino que funciona mal. Y este mal funcionamiento puede derivar bien sea de una de las fases señaladas o de un conjunto de ellas. Para evitar que se produzcan tales sesgos (discriminaciones, parcialidad), proponen algunas medidas que deberían de implementarse conjuntamente:

Como primera medida, permitir inputs en el diseño de los algoritmos de grupos de usuarios con los que abordar estos sesgos. Es decir, que los grupos de usuarios probables (especialmente aquellos que ya están marginados) participen en el diseño de algoritmos y dispositivos como otra forma de garantizar que los sesgos se aborden desde las primeras etapas del desarrollo tecnológico. Como segunda medida, las contramedidas para combatir los prejuicios deben convertirse en la norma del aprendizaje automático. Tanto la primera como la segunda medida propuestas se centran en la fase de recogida de datos, etiquetado de datos y entrenamiento de datos, donde suelen producirse el mayor número de sesgos, como ya se ha explicado. Pero ello no obsta para que en fases ulteriores -control de datos, presentación al cliente- no puedan incluirse nuevos sesgos. De ahí, que, como todo este proceso exige incidir en el control, convenga añadir una tercera medida, que sería la de auditar el algoritmo, de manera que sea capaz de generar confianza. Hay algunas empresas especializadas en esta auditoría algorítmica -como Eticas Consulting, que ha elaborado la primera Guía de Auditoría Algorítmica-, en la que se presta especial atención a los sesgos. La implementación de la IA proseguirá su imparable avance, pero con mayor razón que si se tratara de un producto o servicio cualquiera, necesita generar confianza entre los usuarios.

2.3. ¿Pueden ser justos los algoritmos?

Dada la exigencia de no-discriminación y respeto a la igualdad que se exige a los algoritmos, cabe preguntarse si pueden ser justos y/o equitativos y/o imparciales en el aprendizaje automático (*machine learning*) y, si es así, se debería concluir qué se les debería exigir para ajustarse a tales reivindicaciones. Tal exigencia no deja de ser una utopía ya que, si se traslada ese debate al ámbito real, es conocido el debate sobre la justicia y las dificultades de llegar a un consenso sobre la justicia, como también sobre la equidad y la imparcialidad (*fairness*).

Justo / equitativo / imparcial no son exclusivamente construcciones jurídicas sino esencialmente morales. Aquí surge una primera dificultad: que los algoritmos aprendan nociones morales, que los algoritmos minimicen los daños que causan. Habría que establecer en qué modelos de algoritmos se producen sesgos; habría que medirlos (cada resultado matemático de sesgos responderá a un tipo de sesgo); habría que identificarlos, lo que exigiría establecer y definir qué principios éticos hay detrás de cada definición. Son muchos los interrogantes que subyacen:

i. ¿Cómo se mide el *fairness* en un proceso de decisión? Por un lado, habría que diferenciar el *fairness* de grupo (tratar a dos grupos de manera equivalente) del *fairness* individual (tratar a dos personas de manera similar) (Barocas, Hardt y Narayanan, 2017; Binns, 2018), pero tomando en consideración que la formalización del concepto de *fairness* entre individuos y grupos sociales da lugar a líneas de incompatibilidad y de fricción;

ii. ¿Cómo definir los grupos? ¿Por atributos (género, etnia, orientación sexual o política)?

iii. ¿Con qué porcentaje de equidad nos podemos conformar para considerar que un algoritmo es justo? ¿Un 50%? Para evitar los sesgos, el algoritmo debe tratar de forma similar a dos grupos o dos individuos, pero ¿en qué porcentaje? ¿cuál es el mínimo? En el caso de un sesgo de género, la búsqueda del algoritmo equitativo, ¿podría potencialmente aplicar la forma de "acción positiva" o acción afirmativa, con el propósito de compensar directamente alguna categoría o grupo en desventaja? (Cantero Álvarez, 2021).

iv. Incluso, cabría preguntarse si, en algunos casos, ante la dificultad de lograr un consenso sobre la "justicia algorítmica", podría considerarse lograda esa justicia con la "explicabilidad". Por ejemplo, el algoritmo de una entidad bancaria no concede un crédito y su algoritmo explica el porqué. Es decir, que sea suficiente con que el algoritmo se pueda explicar, aunque no sea justo. Las distintas fases (datos, modelos, decisiones, medición de la causabilidad) deberían ser sometidas a control para verificar la ausencia de parcialidad o de injusticia (Verma y Rubin,2018; Carey, Wu, 2002).

En el caso de que, aun adoptando las precauciones señaladas, se produzcan sesgos, parcialidad y falta de precisión que siempre vaya hacia un mismo lado, hacia un mismo grupo, cabe preguntarse qué se debería de hacer

una vez detectado el sesgo. Como acertadamente se cuestiona Miguel de Berian, ¿Habría que sacrificar la precisión del algoritmo para que acoja por igual a todos los grupos? O bien, ¿habría que utilizar el algoritmo para las comunidades y grupos para los que funciona bien y buscar otras fórmulas para esos otros grupos? Por ejemplo, si un algoritmo funciona bien para la comunidad de los blancos, pero no para los caucásicos, ¿se debería de buscar otro algoritmo para los caucásicos? (De Miguel Beriain, 2021). Por otro lado, hay que recordar que cuando el algoritmo da un diagnóstico y adopta una decisión, al ejecutar un programa, si replica un sesgo, puede que el sistema humano que hay alrededor no se atreva a contradecirlo -habría que argumentar y justificar la adopción de una decisión contraria a la emanada del algoritmo, lo que exige una carga argumentativa mayor-.

3. LOS SESGOS ALGORÍTMICOS Y SU INCIDENCIA EN EL GÉNERO. ALGUNOS CASOS

En una especie de sentimentalismo tecnológico, los algoritmos han sido acusados de ser racistas, sexistas e, incluso, estar estresados (Amato, 2020: 142). Las IA son de diversos tipos (análisis de imágenes, análisis de audio y texto, procesamiento natural del lenguaje, entre otros), cada una con unas capacidades y unos potenciales usos, lo que a su vez da lugar a diferentes tipologías de sesgos.

Los sesgos pueden ser de distinto tipo: de presentación, de filtro, de selección o muestreo, histórico, de agregación, de interacción, etc. (Gutiérrez, 2021). Algunos de estos sesgos actúan juntos y tienen una fuerte influencia en la discriminación. Conviene precisar qué sea el sesgo cognitivo, el sesgo algorítmico y el sesgo algorítmico de género. El cerebro humano utiliza reglas para procesar la información.

El primero, el sesgo cognitivo tiene lugar en el proceso humano de percepción, procesamiento de la información y toma de decisiones. Cuando esas reglas o heurísticas generan desviaciones (errores) respecto de lo que sería la decisión racional de manera sistemática, es cuando surge el sesgo cognitivo. El segundo, el sesgo algorítmico es el que deriva de las decisiones erradas propias de un sistema de IA que provocan o son capaces de provocar un impacto desfavorable respecto de ciertas personas o grupos de personas, que aportan respuestas parciales, sesgadas, con prejuicios, distorsionadas. El problema se presenta cuando estas respuestas afectan de forma importante a los derechos humanos y desembocan en un afianzamiento e incrementos de las brechas existentes. Por último, cuando ese sistema informático propone o toma decisiones erradas para replicar estereotipos de género, se llama sesgo algorítmico de género, que puede derivar en la discriminación algorítmica basada en el género. Aquí confluyen, por una parte, los algoritmos y, por otro,

el estereotipo de género, entendido este como una opinión o un prejuicio generalizado acerca de atributos o características que hombres y mujeres poseen o deberían poseer o de las funciones sociales que ambos desempeñan o deberían desempeñar. "Un estereotipo de género es nocivo cuando limita la capacidad de hombres y mujeres para desarrollar sus facultades personales, realizar una carrera profesional y tomar decisiones acerca de sus vidas y sus proyectos vitales" (Danesi, 2021, p.161).

La relación entre mujer y varón en cuanto al objetivo de la consecución de la igualdad entre ambos géneros ha dado lugar a diversas corrientes y teorías tales como feminismo (Costa Wegsman, 2016), feminismo punitivo (García Figueroa, 2021), giusfeminismo (Cassadei, 2017), y ecofeminismo, entre otras. El término que ha concentrado la atención y es objeto de encendidos debates es el de género (discriminación de género, violencia de género). A diferencia del sexo -que se define como las características biológicas innatas de cada persona **(**fisiológicas, hormonales y genitales) que distinguen a varones y mujeres-, el género es una construcción sociocultural que hace referencia a las cualidades, comportamientos y funciones adscritas socialmente a varones y mujeres de forma diferenciada y jerárquica en función de sus diferencias biológicas.

Los casos de discriminación por razón de sexo consisten en situaciones en las que los hombres o las mujeres reciben un trato menos favorable que las personas del sexo opuesto. La protección contra la discriminación por razón de sexo, en el marco de la protección de datos, ha sido objeto de una profusa normativa al respecto desde el ámbito de la Unión Europea -la Resolución del Parlamento Europeo, de 14 de marzo de 2017, introdujo cautelas referentes a las implicaciones de datos masivos que afectaban a los derechos fundamentales, así como sobre el principio de no-discriminación en Europa- (Venedis/Senden, 2020). Si se consultan los indicadores que miden el nivel de igualdad entre los países de la Unión Europea para saber en qué nivel de consecución de la igualdad nos encontramos, Suecia y Dinamarca encabezan el índice anual que elabora el Instituto Europeo de Igualdad de Género. A la cola están Grecia y Hungría. España se encuentra en el sexto puesto (Álvarez, 2021), situándose por encima de la media europea. El Informe publicado por la Fundación Alternativas, en 2021, titulado 'Inteligencia artificial e igualdad de género. Un análisis comparado entre la UE, Suecia y España', resulta esclarecedor para comprender la incidencia de la IA en el género (Ortiz de Zárate Alcarazo/Guevara Gómez, 2021).

El surgimiento de las plataformas como modelo económico dominante ha aumentado exponencialmente los sesgos. El *machine learning* hace posibles todas las interacciones que tenemos con plataformas como Amazon, Facebook,

Google y Netflix. La IA decide qué ofrecer mediante la personalización de los resultados. La desigualdad en la publicidad, el cine, los videos musicales y la televisión cobra una nueva vida cuando las plataformas toman decisiones algorítmicas basadas en estos contenidos, lo que potencialmente multiplica los prejuicios y establece un círculo vicioso. Redes sociales, plataformas de compra online, motores de búsqueda, chabots (texto) y asistentes de voz (audio), software de reconocimiento facial, son susceptibles de que se les hayan colado sesgos. Actualmente, los sistemas de IA se utilizan en el ámbito laboral (para decidir a quién se contrata); sanitario (calidad del tratamiento médico); bancario (para decidir la concesión de un préstamo bancario); policial (en el sistema predictivo, para decidir si nos convertimos en sospechosos en una investigación policial). Investigadores en la materia han realizado diversas pruebas y ensayos, y han llegado a la conclusión de que queda probado que los algoritmos "tienen prejuicios", entre ellos, los de género. Basta citar algunos casos y estudios que se han realizado para probar los sesgos sexistas.

3.1. Datos sesgados y *machine learning*

Algunos programas de computación trabajan asociando una palabra con otra mediante vectores. Apoyándose en significados semánticos, se establecen analogías y conexiones entre términos en un contexto de *machine learning*.

Un primer ejemplo es el de Google News. Combinando sesgos de interacción, selección y presentación, el sistema de publicidad en línea de Google propone los trabajos mejor remunerados a los hombres. Para demostrarlo, se han realizado diversas pruebas con Google, a través de la técnica de mapeo de palabras, y a partir de datos de Google News (la base más extensa de las que hay). El resultado es que contiene prejuicios y sexismo (Martínez/Matute, 2020; Salas, 2017). Por ejemplo, se puso a prueba un algoritmo:

El hombre es a un rey lo que la mujer es a "X" // X = reina (decía la máquina).

El hombre es a programador informático lo que la mujer es a "X" // X = ama de casa

Google News, en el extremo relacionado con *she* (ella) sitúa profesiones como ama de casa, recepcionista, bibliotecaria, peluquera, niñera, contable, etc; mientras que, en el lado más masculino, en el extremo de *he* (él) figuran términos como profesor, capitán, filósofo, financiero, locutor, mago, jefe, etc. Un estudio, aplicando el denominado Test de Asociación Implícita (TAI), un método comúnmente utilizado para medir los prejuicios en los seres humanos ha demostrado que cuando estos sistemas aprenden un idioma a partir de textos ya existentes se ven contagiados de los mismos prejuicios raciales o de

género incluidos en el lenguaje (Caliskan/Bryson/Narayan, 2020). Los algoritmos de motores de búsqueda en internet han aprendido a asociar mujeres con imágenes de cocinas, basado en decenas de miles de fotografías de internet, porque aparecen más mujeres que hombres fotografiadas en cocinas en la Web (Del Castillo, 2020; Hao, 2021).

Un segundo caso es el del chabot "Tay.AI", lanzado por Microsoft en 2016. Se trataba de una inteligencia artificial que se suponía que debía aprender leyendo tuits e interactuando con otros usuarios de la plataforma Twitter. En su descripción decía: "Cuanto más hablas, Tay se vuelve más lista". Pero con solo unas horas de funcionamiento, Tay empezó a tuitear textos de contenido sexista y racista, convirtiéndose en una máquina generadora de discursos de odio y de discriminación, y tuvo que ser desconectada por Microsoft. La compañía intentó excusarse en que se había producido un ataque informático, pero nunca se pudo demostrar.

Un tercer caso, resultado de una selección discriminatoria producida por técnicas de *data mining* al servicio de la *workforce analitycs*, cada vez más habitual en contextos profesionales, fue el de la IA que utilizaba Amazon, en su proceso de contratación de trabajadores. En 2018 se conoció que la multinacional estadounidense había desechado su herramienta de IA con la que llevaba cuatro años seleccionando a los candidatos para sus puestos de trabajo porque era sexista. Los modelos informáticos de Amazon habían sido entrenados siguiendo los patrones observados en los *curricula* presentados a la empresa durante una década. Pero como el sector tecnológico está altamente masculinizado, la mayoría de los *curricula* que utilizaron para el aprendizaje del automatismo eran de hombres. "Incluso si le dices a los algoritmos de IA que no miren el género, ellos encontrarán otras formas para averiguarlo". El algoritmo de Amazon penalizaba los *curricula* que incluían palabras relativas al género femenino. Aunque no se indicara el sexo en la solicitud, el indicar "campeona de ajedrez en el año x" bastaba para detectar el género (Baeza Yates/Peiró, 2019). Cuando se analizan los riesgos de que la IA genere discriminación, se suele poner el foco en el problema de las "cajas negras" de los algoritmos, como el caso de Amazon citado. De ahí cabría deducir que son preferibles las decisiones adoptadas por seres humanos para evitar las *black box*. Sin embargo, esta conclusión no toma en consideración el hecho de que la mente humana también es, algún sentido, una "caja negra". Y, a diferencia de lo que sucede con los algoritmos, no podemos hacer nada para evitar las "cajas negras humanas" (Tolosa, Dibo, 2021, 172).

Por último, un cuarto caso fue el de Apple Card, que se conoció cuando, en 2019, David Heinemeier Hansson denunció, a través de su cuenta de Twitter, que su Apple Card -tarjeta de crédito emitida por Goldman Sachs- le ofrecía

una línea de crédito veinte veces mayor que la ofrecida a su esposa, a pesar de que ambos presentaban declaraciones de impuestos conjuntas y él tenía peor calificación crediticia. A la reclamación que presentó, los directivos de Apple Card se habían limitado a responder "*It's just the algorithm*".

Posiblemente, en todos estos casos, el problema deriva de la fase de entrenamiento de los algoritmos, es decir, que los datos de los que ha aprendido sean el origen y causa de esos sesgos de género. Los algoritmos de *machine learning* no están sesgados desde su origen, sino que aprenden a ser parciales (Danesi, 2021, 45).

3.2. Softwares de reconocimiento facial

En una cultura de la imagen como en la que nos encontramos, el papel de los datos audiovisuales es el que prima, y contribuye a profundizar y a perpetuar la discriminación hacia la mujer. Los datos audiovisuales así lo confirman, y ello tanto con respecto a la imagen como al sonido.

Los softwares de reconocimiento facial usan bases de datos con etiquetas para identificar el color de la piel, la forma del rostro, el grosor y color del pelo, entre otras características faciales, con distintos objetivos tales como desde diagnósticos y tratamientos médicos, hasta desbloqueo de dispositivos móviles y cajeros automáticos. Además de que su uso afecta a los derechos de privacidad e intimidad -lo que obliga a adoptar cautelas al respecto-, también hay evidencias de que los softwares de reconocimiento facial resultan problemáticos para las mujeres, y ello porque trabajan con bancos de datos de imágenes sesgados. Así, dichos bancos de imágenes ofrecen, mayoritariamente imágenes que reflejan la cultura occidental (y, más concretamente, la anglosajona). De hecho, según ha expuesto un trabajo publicado en la revista *Nature*, especializada en temas de tecnología e IA, más del 45% de los datos de ImageNet -una de las principales referencias en la investigación de visión artificial- proviene de los Estados Unidos. A pesar de que China y la India representen por sí mismas más de un tercio de la humanidad, ambas naciones aportan sólo el 3% de los datos de ImageNet (Schiebinger, 2018; Doshi, 2018; Collett/Dillon, 2019). Además, suele haber una sobrerrepresentación de varones de piel clara y una subrepresentación de personas de piel oscura en general y, concretamente, de mujeres.

Con respecto al contenido visual y fallos en herramientas de reconocimiento facial, se han llevado a cabo varios estudios. Uno de ellos ha versado sobre una comparativa de la distribución de género en las fotos recuperadas por Bing (motor de búsqueda) para la consulta "persona" con respecto a diferentes cualidades. Este estudio concluye que las imágenes de mujeres se vinculan con mayor frecuencia con rasgos cálidos y emocionales,

mientras que los rasgos que indican acción e inteligencia están vinculados preferentemente con fotos de hombres. La imagen del género femenino que ha pervivido durante mucho tiempo ha sido la de empatía, compasión, servilismo y cuidado, y se proyecta en los estereotipos que también asumen los algoritmos. Otro estudio, denominado *Gender Shades,* se ha llevado a cabo sobre tres clasificadores comerciales -Face API de Microsoft, Watson Visual Recognition API de IBM, y Face ++, una compañía con sede en China- y ha puesto de relieve que las mujeres, especialmente las de piel más oscura, son el grupo peor reconocido (Ortiz de Zárate Alcarazo/Guevara Gómez, 2021). Asimismo, los conjuntos de datos pueden exhibir espacios ciegos (ausencia) o puntos críticos (exceso) que terminen en sesgos de género. Por ejemplo, los primeros 112 hallazgos cuando se busca "CEO" (director/a ejecutivo/a) en *Google Images* son, en su mayoría, imágenes de hombres.

3.3. Chabots y asistentes de voz

Los estereotipos de género tienen lugar de diversas formas, bien sea porque las voces de las mujeres en los medios se asocian con ruidos estridentes y desagradables o con funciones dóciles y serviles, bien porque hay sistemas de asistentes de voz "que no reconocen las voces agudas" porque se "han entrenado con voces masculinas", o también porque los historiales que realizan las redes sociales están "basados en estereotipos" y "concepciones previas" que "incentivan la polarización" (Montañés, 2021).

El sexismo también está presente en los chabots y en algunos asistentes de voz que reproducen estereotipos de sumisión y servilismo femeninos. Asimismo, la preferencia de voces femeninas para asistentes digitales puede derivarse de las normas sociales de las mujeres como cuidadoras y otros sesgos de género socialmente construidos que anteceden a la era digital.

Un buen ejemplo de ello es el Informe de la UNESCO, de 2020, titulado *I'd Blush If I Could. Closin gender divides in digital skills trough education* ("Me sonrojaría si pudiera") que deriva su título de la respuesta dada por Siri cuando un usuario le dice: "¡Eres una zorra!". El Informe analiza cómo los asistentes de voz de IA, Alexa de Amazon, Siri de Apple y Cortana de Microsoft, que usan las voces de mujeres jóvenes, propagan los prejuicios de género. Incluso el asistente virtual en el smarthphone y dispositivos de casa sólo tienen voz femenina, al menos en español. Ello viene a reforzar el papel de "asistencia" que se atribuye al género femenino, a la vez que sus respuestas a mensajes abusivos son tibias en lugar de cortantes, lo que consolida la imagen de subordinación de la mujer- como se refleja en la "Entrevista con el robot Sophia", de Hanson Robotics.

Hay diversos retos para abordar el contenido audiovisual sesgado. Hay fallas en los datos, que se constata en tres desafíos principales: falta de claridad sobre la presencia de datos; falta de comprensión sobre cómo funciona el aprendizaje automático (ML) -hay que subrayar que los "Principios de aprendizaje automático responsable" incluyen la "evaluación de sesgos" como uno de sus ocho principios-; y falta de incentivos para que las corporaciones prevengan y corrijan los sesgos. Estos desafíos están presentes en todos los procesos algorítmicos, no solo los que afectan a las mujeres.

El problema es que los sesgos algorítmicos se encuentran en todas las plataformas. En una sociedad en la que todo lo que hacemos se transforma en datos, procesados, digeridos y mediados por algoritmos que contribuyen a decisiones críticas, los derechos de las mujeres dependen de una ética y una agenda de investigación que promueva la igualdad de manera proactiva (Rodríguez Martínez, 2020; Ricoy Casas, 2021). Tales penalizaciones suelen pasar inadvertidas, pero acaban ahondando en esa brecha de género. ¿Qué pasa con todos los procesos que ya están mecanizados y desconocemos cómo nos afectan? ¿Cómo sabrá una mujer que se la privó de ver un anuncio de trabajo? ¿Cómo podría una comunidad pobre saber que está siendo acosada policialmente por un software? ¿Cómo se defiende un delincuente de una minoría étnica que ignora que un algoritmo le señala? ¿Cómo neutralizar el riesgo de no poder determinar responsabilidades y retrotraer los efectos de las decisiones tomadas por sistemas de IA?

La propia tecnología sigue siendo la que, en su evolución, requiere de respuestas jurídicas a nuevas amenazas a derechos fundamentales y a nuevas formas de comisión de delitos. El metaverso, diseñado por el creador de Facebook, Marx Zuckerberg, y en el que han confluido todas las tecnologías, es un ejemplo de ello. Desde el momento de su creación ha sido una fuente de problemas jurídicos. La compra de terreno virtual y la instalación de marcas comerciales reconocidas plantea cuestiones que atañen a la seguridad jurídica de empresas que hacen negocios en el metaverso; protección de los consumidores; propiedad / posesión; jurisdicción aplicable; responsabilidad civil y penal -se ha denunciado un primer caso de acoso sexual virtual en *Horizon Worlds* por el que un avatar había sido tocado de forma no consentida y con intención sexual por otro avatar- (Jiménez de Luis, 2021) así como la vulneración de derechos fundamentales -como podría ser el derecho al honor-.

4. PROPUESTAS PARA UNA INTELIGENCIA ARTIFICIAL PROACTIVA EN LA LUCHA CONTRA LOS SESGOS DE GÉNERO EN PARTICULAR

Con el ánimo de poder formular algunas propuestas que contribuyan no sólo a luchar contra los sesgos algorítmicos de género sino también y fundamentalmente, a impulsar una línea de trabajo proactiva, sirviéndose de la

propia IA para lograr la igualdad (lo cual no implica necesariamente un enfoque de género, sino de lograr una igualdad formal y material como mínimo) conviene revisar en el Derecho comparado el estado de la cuestión para, después, ofrecer algunas áreas de trabajo en las que el uso de la IA favorecerá la consecución de la igualdad de género.

4.1. Una mirada al ámbito internacional

Una mirada al ámbito internacional permite testar si se ha generado una conciencia sobre la problemática de los sesgos de género en algunas aplicaciones de IA. Para empezar, son pocos los textos e Informes de ámbito internacional que defienden e incorporan una perspectiva de género a la robótica e Inteligencia artificial (Wajcman, 2004; Bray, 2007; Huertas Sánchez, 2021). Son escasas las Declaraciones y textos sobre IA que incluyen tal enfoque de género (Deva, 2020): *The Montreal Declaration for the Responsible Development of Artificial Intelligence* no se refiere de forma explícita a la integración de una perspectiva de género; el Marco ético de AI4People para una buena sociedad de IA (Floridi) se limita a mencionar la diversidad/género una vez. Tanto la Recomendación del Consejo de la OCDE sobre la IA (*Recommendation of the Council on Artificial Intelligence*), de 2019, como los Principios de IA del G20 (G20 *Ministerial Statement on Trade and Digital Economy*), subrayan la importancia de que la IA contribuya a reducir la desigualdad de género, pero no especifican cómo podría lograrse.

Los Estados miembros de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura -UNESCO- son los que han expresado una mayor sensibilidad con respecto al género en varios de sus Informes -como el *Artificial intelligence and gender equality: key findings of UNESCO's Global Dialogue*, de 2020-. Especial mención conviene hacer al "Proyecto de texto de la Recomendación sobre la ética de la Inteligencia Artificial", de la UNESCO, de 2021. Además de que se trata de la primera norma mundial sobre la ética de la inteligencia artificial, su ámbito de actuación número seis versa sobre "género", enumerando un conjunto de deberes que competen a los Estados Miembros para lograr la igualdad de género en los sistemas de IA. Específicamente y con relación a los sesgos, en el punto 90 se dice: […] velar por que los estereotipos de género y los sesgos discriminatorios no se trasladen a los sistemas de IA, sino que se detecten y corrijan de manera proactiva. […]. El texto tiene la virtualidad de que no se limita a tratar el género como un problema de sesgos al que hay que dar una solución, sino que va más allá, estableciendo como deber de los Estados que las normas relativas a la IA y a la automatización, de una manera transversal, tengan un enfoque de género.

El reto no es exclusivamente el de detectar y eliminar los sesgos algorítmicos sexistas sino aprovechar el potencial que brinda la IA para

promover un enfoque de género, impulsando la efectiva igualdad de género-. El citado Informe de la Fundación Alternativas analiza y describe la experiencia de Suecia y su apuesta por la igualdad de género puede ser un modelo de referencia. Suecia -primer país de la UE en igualdad de género- utiliza la IA como herramienta para poder identificar situaciones de discriminación hacia las mujeres y, así, proponer soluciones para mejorar su inclusión.

Entre las iniciativas, destaco las que se han desarrollado en los siguientes ámbitos: en el socio-jurídico, en el marco del proyecto Ceretai de la Agencia Sueca de Innovación (Vinnova) se desarrollan herramientas automatizadas y diseñadas para la detección de normas, patrones y estereotipos discriminatorios presentes en la cultura popular, con el fin de cuantificar y visibilizar comportamientos discriminatorios que sufren las mujeres, y que a menudo pasan inadvertidos. En la misma línea, la iniciativa NoBias utiliza el *machine learning* para eliminar los prejuicios y las normas a través de explicaciones que muestran usos no inclusivos y discriminatorios del lenguaje, con el fin de difundir el lenguaje inclusivo en el mundo de la empresa. En el ámbito de la educación, Suecia ha planteado utilizar la IA para evaluar tareas y exámenes de modo que se eviten sesgos por género (proyecto Gradecam), y también para detectar casos de *bullying* desde momentos tempranos y para evitar sesgos de género en los procesos de admisión en las Universidades. En el ámbito laboral, con la finalidad de promover la igualdad económica entre hombres y mujeres, se persigue analizar las bases de datos de las empresas para identificar desigualdades en el salario entre hombres y mujeres. Por su parte, el proyecto Rikare II busca reducir los sesgos de género en la concesión de financiación para proyectos emprendedores de áreas tradicionalmente masculinas, a través de un algoritmo. En último lugar, en cuanto a la lucha contra la violencia de género, la IA se utiliza para detectar comportamientos agresivos o violentos hacia las mujeres, analizando llamadas de emergencia, imágenes de vídeo o publicaciones en redes sociales (proyectos Grace Health). Por último, habría otros campos, como el disfrute igualitario de salud (proyectoy Bonzun) (Rodríguez Canfranc, 2021).

El Proyecto de *Déclaration européenne sur les droits et principes numériques pour la décennie numérique*, aprobado el 26 de enero de 2022, en su Preámbulo, recuerda que la estrategia de la UE en materia de derechos digitales es, entre otras exigencias, plenamente respetuosa con los derechos fundamentales, con la protección de datos, y con la no-discriminación. Explicita que: "les valeurs de l'Union et les droits des personnes reconnus par le droit de l'Union soient respectés tant en ligne qu'hors ligne", para alejar la sensación de impunidad que parece permear el ámbito digital.

Por su parte, España ha remitido en 2020, su Estrategia para la IA, en la que, siguiendo las orientaciones de Informes y Documentos de la Unión Europea, ya incluye referencias explícitas e implícitas al género -el desafío social 1 es "Reducir la brecha de género del ámbito de la IA en empleo y liderazgo": "[…] la igualdad de género ha de ser uno de los objetivos transversales de la presente estrategia"- lo que ya constituye un adecuado punto de partida para transitar hacia una donde el género no sea un factor de discriminación.

La *Carta de derechos digitales*, adoptada en julio de 2021 por el Gobierno de España, y que se inscribe en el contexto de la Estrategia Española Nacional de Inteligencia Artificial -ENIA- de 2020, aunque carece de efectos normativos, en su punto seis, apuesta por "Establecer un marco ético y normativo que refuerce la protección de los derechos individuales y colectivos, a efectos de garantizar la inclusión y el bienestar social". Aunque la Carta no tiene carácter normativo, se presenta como la línea a seguir en la legislación y políticas públicas. La Carta hace especial incidencia con respecto a los derechos en entornos específicos, como es el caso los derechos ante la Inteligencia artificial. En el Capítulo XXV, titulado "Derechos ante la inteligencia artificial", en el apartado primero, defiende un enfoque centrado en la persona y en su inalienable dignidad; en el apartado segundo, con respecto al desarrollo y ciclo de vida de los sistemas de inteligencia artificial, en su apartado a) preceptúa el derecho a la no discriminación: "Se deberá garantizar el derecho a la no discriminación cualquiera que fuera su origen, causa o naturaleza, en relación con las decisiones, uso de datos y procesos basados en inteligencia Artificial". Si la finalidad planteada es clara, el problema radica en el cómo conseguir esa no discriminación. De ahí que, continúa en el apartado b: "Se establecerán condiciones de transparencia, auditabilidad, explicabilidad, trazabilidad, supervisión humana y gobernanza. En todo caso, la información facilitada deberá ser accesible y comprensible". Y se completa con lo establecido en el apartado tercero, según el cual "Las personas tienen derecho a solicitar una supervisión e intervención humana y a impugnar las decisiones automatizadas tomadas por sistemas de inteligencia artificial que produzcan efectos en su esfera personal y patrimonial".

4.2. Cinco áreas de trabajo para promover la igualdad de género a través de la IA

El uso de la IA como herramienta para promover la igualdad de género se presenta como un reto que se podría desarrollar a través cinco áreas de trabajo, a las que hago seguidamente referencia, siguiendo las propuestas del Documento de la Fundación Alternativas:

- La educación, incentivando desde niñas el interés de las menores por las materias STEM (científicas y tecnológicas como las

Matemáticas, Ingenierías, Ciencias y Tecnología) (Ricoy Casas, 2021; Garrido Gómez, 2020). Parte de la problemática existente en el diseño de algoritmos se debe a que menos del 25% de las personas que se dedican a la investigación en IA son mujeres. Como explica el citado Informe 'Inteligencia artificial e igualdad de género. Un análisis comparado entre la UE, Suecia y España', las mujeres solo representan el 12% de los autores de artículos en las principales conferencias sobre *machine learning*, y el 13,83% de los miles que se escriben en general sobre Inteligencia Artificial. Además, es muy escasa la presencia de mujeres en los grupos y equipos que diseñan y desarrollan IA. Para ponerlo de relieve, *Women in AI Ehtics*, proyecto social que arroja luz sobre los riesgos que los algoritmos sesgados pueden traer a las vidas humanas, individual y colectivamente, en particular a la discriminación de mujeres y grupos de personas más tradicionalmente marginados, publica anualmente la lista de las 100 mujeres brillantes en la ética de la IA.

- La elaboración de guías para la implementación de la IA en las Administraciones públicas (AAPP) con una perspectiva de género.
- La lucha contra la violencia de género, por ejemplo, a través del desarrollo de *chatbots* que faciliten los procesos administrativos para denunciar y que presten asistencia en tiempo real a las víctimas. En España se valora adecuadamente "El Sistema de Seguimiento Integral en los casos de Violencia de Género" (Sistema VioGén). Se trata de un modelo desarrollado por el Ministerio del Interior en 2007 que, mediante el aprendizaje algorítmico, evalúa entre otras cuestiones, el riesgo de que el mismo agresor vuelva a actuar, haciendo así una función preventiva y de protección para la víctima. A partir de la información de las denuncias interpuestas por mujeres y otros indicadores, valora el riesgo de que una mujer vuelva a sufrir una agresión, de manera que ofrece información para tomar decisiones policiales que permitan garantizar la protección de las mujeres víctimas de violencia.

Desde la Fundación Éticas y en colaboración con la Fundación Ana Bella, se ha comenzado un estudio que explora el impacto e imparcialidad de este algoritmo, especialmente en las mujeres más vulnerables. El estudio cuenta con tres objetivos principales: i) Examinar si el sistema hace diferencias a niveles predictivos según el grupo de mujeres al que se aplique (migrantes, españolas,

jóvenes, con hijos a su cargo, etc.); ii) Tener en consideración la experiencia y opinión basadas en el uso de esta herramienta de víctimas, sus representantes legales y asociaciones que luchan contra la violencia de género; iii) Desarrollar métodos para auditar externamente el sistema automatizado de evaluación de riesgos de VioGén en ausencia de datos administrativos o acceso al código, para ofrecer guías a aquellas organizaciones sin acceso al mismo.

De nuevo subrayo que la decisión automatizada debe ser supervisada por un ser humano, lo que viene no sólo preceptuado por la normativa reguladora sino por la propia experiencia. Basta recordar que, con fecha 30 de septiembre de 2020, la Sala de lo Contencioso-administrativo de la Audiencia Nacional ha condenado al Ministerio del Interior por la deficiente protección que la Guardia Civil otorgó a una mujer que solicitó una orden de protección. Un cuestionario de cribado calificó su situación como de "riesgo bajo". Sin realizar más averiguaciones los agentes calificaron el riesgo como "no apreciado". Y esta misma valoración fue determinante para que también el juzgado denegase la medida de protección a la fallecida. El cuestionario de cribado de IA no apreció un riesgo que, con una entrevista personal y aplicando la perspectiva de género, se habría considerado como de riesgo, y se habría dado la debida protección a la víctima, habiéndose podido evitar su muerte. La respuesta se limitó a la recogida de datos automatizados, pero no previno la violencia ni reevaluó el riesgo por medio de agentes especializados en su tratamiento y sensibilizados con la lacra de la violencia de género. La sala reconoce el quebrantamiento de la obligación estatal positiva de proteger. Existían indicios: su marido tenía antecedentes por maltrato en el país de origen, pero no se comprobaron; la violencia se ejercía delante de los menores e incluso delante de la madre del agresor, pero no se les tomó declaración. Las trabajadoras sociales describen una víctima totalmente sometida, con pánico a su agresor. Pero nada de esto fue suficiente para reevaluar el riesgo que un sistema de IA otorgó de manera errónea. Precisamente aquí es donde surge la responsabilidad estatal: no se puede depositar en la máquina la responsabilidad de la toma de decisiones que debe hacer el humano (Borges Blázquez, 2020: 66-67).

- La reducción de la brecha salarial y la desigualdad económica entre hombres y mujeres a través del desarrollo de sistemas de IA que contribuyan a favorecer la transparencia retributiva en las empresas (Ramírez, 2021; Romero/Sánchez, 2021). La vigilancia de sesgos de género va más allá de una" política feminista", en la medida en que influye directamente en la igualdad y, en este caso, en el derecho al trabajo. El pasado 3 de diciembre de 2021, el Consejo de Ministros de España ha aprobado, en primera lectura, la nueva ley del empleo que, siguiendo la estela de la estrategia de activación del empleo, aspira a llevar a cabo una transformación radical de los servicios públicos de empleo. Entre las diversas medidas contempladas, destaca el impulso de "la

elaboración de un perfil individualizado del usuario que permita, a través de evidencias estadísticas, mejorar su empleabilidad". Dicho análisis se desarrollará mediante "el uso de la inteligencia artificial y la digitalización de las administraciones públicas, de manera que la futura Agencia Española para el Empleo será la encargada de cruzar estos datos y de garantizar que los algoritmos y las herramientas digitales de tratamiento masivo de datos no penalizan a los demandantes de empleo y no introducen sesgos (existen ejemplos de herramientas algorítmicas que han 'aprendido' e interiorizado sesgos de género, por ejemplo, penalizando a las mujeres)".

Además, habría que incorporar más mujeres- claro está, a quienes tengan competencia para ello y no sólo por una cuestión de cuota de género- a las posiciones ejecutivas en empresas tecnológicas dedicadas a desarrollar IA. Un estudio realizado en 2017 por Emerj demostró que sólo el 13% de las altas posiciones ejecutivas en empresas tecnológicas dedicadas a la IA son ocupadas por mujeres, y, en lo que se refiere al subámbito del lenguaje natural, el porcentaje cae hasta el 5% (Rodríguez Canfranc, 2021).

- La atención a las necesidades de salud específicas de las mujeres a través del desarrollo y uso de sistemas de IA especializados en el diagnóstico de enfermedades que sufren las mujeres como el cáncer de cuello de útero, el cáncer de mama, etc. Se trataría de diseñar e implementar un *chatbot* para ayudar a las mujeres a resolver sus dudas sobre salud.

Las cinco áreas citadas son muy amplias (educación, Administración pública, violencia de género, ámbito laboral, y sistema de salud), y abarcan buena parte de las políticas públicas de un Estado social. En todas ellas, la IA ya ha irrumpido con fuerza y se están llevando a cabo investigaciones sobre las limitaciones de la IA en cada campo, pero también, del potencial que puede aportar en cada una de las Áreas. Por ello, detectar los sesgos de género en fase temprana y neutralizarlos, evitará decisiones erróneas e injustas, a la vez que contribuirá a reforzar la seguridad jurídica.

5. CONCLUSIONES

La "condición algorítmica" está incidiendo en los derechos humanos, que prescriben la igualdad y prohíben la discriminación -entre otras condiciones, por razón de sexo biológico y de género-. En una sociedad en la que todo lo que hacemos se transforma en datos, procesados, digeridos y mediados por algoritmos que contribuyen a decisiones críticas, los derechos dependen de cómo se regulen estos avances. Hay que evitar crear un círculo perverso de discriminación en línea y en la vida real. El sistema jurídico es el responsable de evitar la desigualdad, la parcialidad y la discriminación. Se trata de redefinir el ámbito de lo jurídico (no del Derecho), y la teoría de los derechos humanos se

enfrenta a una auténtica revolución. La elaboración de perfiles, las dificultades de explicabilidad y causalidad de los algoritmos, consolidan a su vez, la dificultad de lograr el diseño de algoritmos "equitativos e imparciales". La normativa podría exigir, por ejemplo, que los sistemas algorítmicos utilizados en el sector público se desarrollen de forma que permitan, como mínimo, la auditoría y la explicación. Lo expuesto en este estudio permite apuntar varias reflexiones finales.

Se ha avanzado en diversos proyectos para minimizar sesgos -como las herramientas que ha desarrollado IBM en los últimos años, tales como AI Fairness 360, AI FactSheets, IBM Watson OpenScale, así como las nuevas capacidades de IBM Watson diseñadas para ayudar a las empresas a crear una IA confiable-. Junto a los avances informáticos, hay que construir diversos frentes (educativos, investigadores) y desde distintos ámbitos (empresas, usuarios, juristas, ingenieros informáticos) para seguir avanzando en detectar y minimizar los sesgos en general, y los de género en particular. Al igual que en la legislación y políticas públicas se trabaja desde un enfoque de equidad, tal perspectiva puede muy bien aplicarse al espacio digital, de manera que todas las fases de implementación de la IA estén orientadas por una equidad de género, desde su diseño y desarrollo hasta su implementación. De lo contrario, hay un riesgo de amplificación de sesgos y su perpetuación con la aplicación de la Inteligencia artificial en la población. A este respecto, la labor de los tribunales también será esencial para controlar los sesgos algorítmicos. Basta recordar el modo en el que el Tribunal Europeo de Derechos Humanos y el Tribunal de Justicia de la Unión Europea han abordado la tecnología. Los tribunales han lanzado un mensaje claro de garantía de la tecnología frente a los derechos fundamentales y de aplicación del modelo europeo frente a quienes se amparaban en el subterfugio de la ausencia de normas o la extraterritorialidad.

Lo que aquí he defendido no es el de fomentar una IA "feminista" guiada por un discurso demagógico. Como jurista, hay que poner todos los medios para evitar que los programas de IA contengan sesgos de cualquier tipo; pero hay que ir más allá, y aprovechar el potencial de la IA para que se pueda convertir en un instrumento de inclusión. Por ejemplo, debe incrementarse la participación de mujeres competentes en programas de diseño y desarrollo de IA, lo que, a la vez exige previamente, un impulso de una *vis atractiva* hacia las materias tecnológicas por parte de las niñas. La educación desde una fase temprana se presenta como una poderosa herramienta de cambio para afianzar el valor y derecho de la igualdad entre sexos.

Hay que ser realistas y reconocer que es imposible eliminar totalmente los sesgos, ya que, bien sea el factor humano, el tecnológico o el social, estarán siempre presentes. Es más, una absoluta eliminación de la diferencia podría

originar un problema de discriminación y desigualdad también. Sin embargo, se debe intentar minimizarlos. Es decir, hay que trabajar para mejorar esa detección de los sesgos e implantar las medidas para neutralizarlos, y ello, a lo largo de todo el ciclo de vida de la IA. Se deben adoptar tanto medidas de índole política y legislativa, como medidas propiamente tecnológicas, con una IA que garantice el escrupuloso respeto de los derechos fundamentales y que resulte "fiable" -fiabilidad que exigen documentos y Directrices éticas de diversos organismos e instituciones, como la UNESCO y la Unión Europea-. Los principios rectores sobre la IA de alto riesgo -el ámbito jurídico lo es en cuanto afecta a derechos fundamentales- requiere transparencia, rendición de cuentas, supervisión humana y ausencia de sesgo y de discriminación. Para ello, los algoritmos deben ser auditables, transparentes y explicables, tal y como se exige en varios documentos internacionales e instrumentos normativos, lo que a su vez requiere establecer un sistema de control y de auditoría. No basta con establecer cómo deben ser los algoritmos sino verificar que, efectivamente, se están cumpliendo los requisitos establecidos.

Por último, se abre una nueva vía de lucha contra los sesgos algorítmicos a partir de la propuesta de algunos neurocientíficos, como Rafael Yuste, con respecto a la conveniencia de crear nuevos derechos humanos, como los neuroderechos, entre los cuales está el "derecho a la protección contra los sesgos algorítmicos", es decir que los conocimientos adquiridos con la neurociencia no establezcan trato discriminatorio ni distinción por razón de raza, color, sexo, idioma, religión, opinión, origen nacional o social, posición económica, nacimiento o cualquier otra condición. Ello requerirá reflexionar sobre si aporta realmente algo nuevo al clásico derecho fundamental de no discriminación (Ienca y Andorno, 2017; Ienca, 2021; Yuste *et al.*, 2021) a la vez que pone el foco de atención en uno de los grandes retos que tiene ante sí la IA: cómo neutralizar los sesgos.

6. BIBLIOGRAFÍA

AA. VV., "Inteligencia Artificial y mujeres, una historia de discriminación", en: *Eticas*. https://www.eticasconsulting.com/inteligencia-artificial-y-mujeres-una-historia-de-discriminacion/

Álvarez, Pilar (2021), "España escala hasta el sexto puesto en igualdad entre hombres y mujeres en la UE", en: *Diario El País*, Madrid (28.10.2021) https://elpais.com/sociedad/2021-10-28/espana-escala-hasta-el-sexto-puesto-en-igualdad-entre-hombres-y-mujeres-en-la-ue.html

Amato, Salvatore (2020), Emozioni sintetiche e sortilegi al silicio", en: *Ars Interpretandi. Rivista di Ermeneutica Giuridica, Algoritmi ed esperienza giuridica*, X, 1, Carocci Editore, Roma, 129-141.

Baeza Yates, Ricardo, Karma Peiró (2019), "¿Por qué la inteligencia artificial discrimina a las mujeres?", en: *médium.com* (22.10.2019) https://medium.com/think-by-shifta/por-qu%C3%A9-la-inteligencia-artificial-discrimina-a-las-mujeres-18b123ecca4c

Barocas, Solon; Moritz Hardt, Arvind Narayann (2017), *Fairness and machine learning. Limitations and Opportunities*. https://fairmlbook.org/pdf/fairmlbook.pdf

Barrio Andrés, Moisés (2020) "Retos y desafíos del Estado algorítmico de Derecho". *Análisis del Real Instituto Elcano* (ARI), Nº. 82, pp. 1-6. https://www.realinstitutoelcano.org/analisis/retos-y-desafios-del-estado-algoritmico-de-derecho/

Beck, Ulrich (2017), *La metamorfosis del mundo*, Paidós Barcelona.

Bellver Capella, Vicente (2021), "Combatir la discriminación, esencia de los derechos humanos. Presentación", en: Bellver Capella, Vicente, Ángeles Solanes Corella (dirs.), *Derechos Humanos y lucha contra la discriminación*: Actas del IV Congreso Internacional sobre Derechos Humanos: celebrado online los días 4 y 5 de febrero de 2021, 7-8.

Bernheim, Aude, Flora Vincent (2019), *L'intelligence artificielle, pas sans elles!*, Belin, Paris.

Binns, Reuben (2018), "Fairness in machine learning: Lessons from political philosophy", en: *Proceedings of Machine Learning Research*, 81, 149-159.

Borges Blázquez, Raquel (2020), "El sesgo de la máquina en la toma de decisiones en el proceso penal", en: *Ius et Scientia*, 6, 2, 54-71.

Bray, Francesca (2007), "Gender and Technology", en: *Annual Review of Anthropology*, 37-52.

Caliskan, Aylin, Joanna J. Bryson, Arvind Narayan (2017), "Semantics derived automatically from language corpora contain human-like biases", en: *Science*, 356, 6334, 136-186 (14.04.2017) https://www.science.org/doi/full/10.1126/science.aal4230

Cantero Álvarez, Héctor (2021), *Justicia algorítmica: el sesgo de los algoritmos*. Proyecto Fin de Carrera / Trabajo Fin de Grado, E.T.S.I. de Sistemas Informáticos (UPM), Madrid. https://oa.upm.es/69093/

Carey, Alycia N, Xinato Wu (2002), *Fairness Field Guide: Perspectives from Social and Formal Sciences.* https://arxiv.org/pdf/2201.05216.pdf

Cassadei, Thomas (ed.) (2015), *Donne, diritto, diritti. Prospettive del giusfemminismo,* Torino, Giappichelli.

Collett, Clementine, Sarah Dillon (2019), *AI and gender. Four Proposals for Future Research. Centre for the Future of Intelligence*, University of Cambrigde, The Leverhulme Centre for the Future of Intelligence. https://www.repository.cam.ac.uk/handle/1810/294360

Costa Wegsman, Malena (2016), *Feminismos Jurídicos*, Ediciones Didot, Buenos Aires.

Danesi, Cecilia Celeste (2020), "Inteligencia Artificial y Derecho", en: *Inteligencia Artificial, Tecnologías emergentes y Derecho. Reflexiones interdisciplinarias*, 1, 39-84.

— (2021), "Sesgos algorítmicos de género con identidad iberoamericana: las técnicas de reconocimiento facial en la mira", en: *Revista Derecho de Familia*, 100, 159-168.

Del Castillo, Carlos (2020), "Si es hombre lleva un martillo, pero si es mujer es un secador: así actúan los sesgos de la Inteligencia Artificial", en: *El Diario.es*. (10.09.2020). https://www.eldiario.es/tecnologia/si-hombre-lleva-martillo-si-mujer-secador-actuan-sesgos-inteligencia-artificial_1_6210120.html

Deva, Surya (2020), *Afrontar el sesgo de género en la Inteligencia Artificial y la automatización*. 10.04.2020. https://www.openglobalrights.org/addressing-gender-bias-in-artificial-intelligence-and-automation/?lang=Spanish

Doshi, Tulsee (2018), "Introducing the Inclusive Images Competition", en: *Blog Google*. (06.09.2018) https://ai.googleblog.com/2018/09/introducing-inclusive-images-competition.html

Fernández, Ana (2019), "Inteligencia artificial en los servicios financieros", en: *Boletín Económico. Artículos Analíticos*, Banco de España, 2, 1-10.

Fernández, C. B. (2020), "Primera sentencia europea que declara ilegal un algoritmo de evaluación de características personales de los ciudadanos", *Wolters Kluwer*, 13 de febrero de 2020. https://diariolaley.laleynext.es/Content/Documento.aspx?params=H4sIAAAAAAAEAMtMSbH1czUwMDAyNDa3NDJUK0stKs7Mz7M1MjACC6rl5aekhrg425bmpaSmZealpoCUZKZVuuQnh1QWpNqmJeYUp6qlJuXnZ6OYFA8zAQCfSdkrYwAAAA==WKE

Fico, Antonio (2018), "Cina: la reputazione del 'buon cittadino' disegnata dai big data", 1 de mayo de 2018. https://altreconomia.it/cina-big-data/

Floridi, Luciano *et al.* (2019), "AI4People's Ethical Framework for a Good AI Society report: opportunities, risk, principles, and recommendations", en: *AI4 PEPOPLE*. https://www.eismd.eu/wp-content/uploads/2019/11/AI4People%E2%80%99s-Ethical-Framework-for-a-Good-AI-Society_compressed.pdf

García Figueroa, Alfonso J. (2021), "La génesis populista del feminismo punitivo", en: *Anales de la Cátedra Francisco Suárez*. Protocolo I. Crisis del Derecho Penal del Estado de Derecho: Manifestaciones y Tendencias, 1, 15-41.

Garrido Gómez, Mª Isabel (2020), "Por una mayor visibilización de la mujer en la educación superior", en: *Quaderns digitals: Revista de Nuevas Tecnologías y Sociedad,* 91, 114-139.

Gutiérrez, Miren (2021), "Ética digital", en: *eldiario.es* (07.02.2021) https://www.eldiario.es/tecnologia/sesgos-genero-algoritmos-circulo-perverso-discriminacion-linea-vida-real_129_7198975.html

Hao, Karen (2019). "Cómo acabar con los algoritmos sexistas que conceden créditos" [traducido por Ana Milutinovic], en: *MIT Technology Review* (27.11.2019). https://www.technologyreview.es//s/11630/como-acabar-con-los-algoritmos-sexistas-que-conceden-creditos

— (2021). "Internet está tan sesgado que, para la IA, las mujeres solo llevan bikini" (03.02.2021) https://www.technologyreview.es//s/13117/internet-esta-tan-sesgado-que-para-la-ia-las-mujeres-solo-llevan-bikini

Harari, Yuval Noah (2016), *Homo Deus: Breve historia del mañana*, Debate, Barcelona.

Herranz, Arantxa (2019), Xataca.com (28.11.2019) https://www.xataka.com/robotica-e-ia/que-seran-capaces-inteligencia-artificial-machine-learning-10-anos-mayores-expertos-nos-responden

Holland, Sarah, Ahmed Hosny, Sarah Newman, Joshua Joseph, Kasia Chnmielinski (2018), "The Dataset Nutrition Label: A Framework To Drive Higher Data Quality Standards", en: *Databases* (cs.DB); Computers and Society (cs.CY). Cornell University (09.05.2018). arXiv:1805.03677

Huertas Sánchez, María Antonia (2021), **"**¿Por qué es necesario incorporar visión de género a la robótica y la inteligencia artificial?", en: *The Conversation***,** Barcelona: UOC - Universitat Oberta de Catalunya (17.06.2021) https://theconversation.com/por-que-es-necesario-incorporar-vision-de-genero-a-la-robotica-y-la-inteligencia-artificial-160655

Ienca, Marcello (2021), "On neurorights", en: *Frontiers in Human Neurosciencie*, 15, 1-11. https://doi.org/10.3389/fnhum.2021.701258

Ienca, Marcello y Andorno, Roberto (2017), "A new category of human rights: neurorights", en: *Research in Progress Blog*. http://blogs.biomedcentral.com/bmcblog/2017/04/26/new-category-human-rights-neurorights/

Jiménez de Luis, Ángel (2021), "Denuncian un primer caso de acoso sexual virtual en el metaverso de Facebook", en: *Diario El Mundo*, 21 de diciembre de 2021.

Kleinberg, Jhon, Jens Ludwig, Sendhil Mullainathan, Cass R Sunstein (2018), "Discrimination in the Age of Algorithms", en: *Journal of Legal Analysis*, 10, 113-174.

Lettieri, Nicola (2020), *Antigone e gli algoritmi. Appunti per un approccio giusfilosofico. Prassi sociale e teoria giuridica*, Stem Mucchi Editore, Modena.

Lasalle, José Mª (2019). *Ciberleviatán: El colapso de la democracia liberal frente a la revolución digital*, Arpa Editores, Madrid.

Llano Alonso, Fernando Higinio (2018), *Homo Excelsior. Los límites ético-jurídicos del transhumanismo,* Tirant lo Blanch, Valencia.

Martínez, Narora, Helena Matute (2020), *El sexismo en los algoritmos: una discriminación subestimada,* Bilbao: Universidad de Deusto **(**22.06.2020) https://theconversation.com/el-sexismo-en-los-algoritmos-una-discriminacion-subestimada-140790

Menéndez Sebastián, Eva M.ª (2021), "Buena administración, algoritmos y perspectiva de género", en: Bonorino Ramírez, Pablo Raúl, Patricia Valcárcel Fernández, Rafael Fernández Acevedo (coords.), *Nuevas normatividades. Inteligencia artificial, derecho y género,* Thomson Reuters Aranzadi, Navarra, Cizur Menor, 35-62.

Mercader Uguina, Jesús R. (2021) "Editorial. Discriminación algorítmica y derecho granular: nuevos retos para la igualdad en la era del big data", en: *Labos, Revista de Derecho del Trabajo y Protección social*, 2, 2, 4-10.

Merino, Marcos (2018), **"**Hay quien critica a mucha inteligencia artificial como racista/etnocéntrica, pero el problema está en los datos", en: *Xataca.com* (18.12.2018) https://www.xataka.com/robotica-e-ia/hay-quien-critica-a-mucha-inteligencia-artificial-como-racista-etnocentrica-problema-esta-datos

— (2019), "El problema de esta IA de Google no gira en torno a ningún 'sesgo racial', sino a una polémica socio-política", en**:** *Xataca.com* (13.08.2019). https://www.xataka.com/inteligencia-artificial/problema-esta-ia-google-no-gira-torno-a-ningun-sesgo-racial-sino-a-polemica-socio-politica

Montañés, Erika (2021), "La inteligencia artificial también es machista: Igualdad combate ahora los algoritmos con sesgo de género", en: Diario *ABC***,** Madrid (23.09.2021)

Myers West, Sarah, Merediht Whittaker, Kate Crawford (2019), "Discriminating Systems. Gender, Race, and Power in AI", en: *AI Now Institute* (AI NOW) https://ainowinstitute.org/discriminatingsystems.html.

Ortiz de Zárate Alcarazo, Lucía; Guevara Gómez, Ariana. (2021), "Una inteligencia artificial feminista es posible (y necesaria)", en: *Diario El País*, Madrid (06.07.2021)

— (2021). Informe 'Inteligencia artificial e igualdad de género. Un análisis comparado entre la UE, Suecia y España'. Fundación Alternativas (10.06.2021).

Pérez Luño, Antonio Enrique (2021), "El posthumanismo no es un humanismo", en: *Derechos y libertades* 45, 17-40.

Ramírez, Helena (2021a), "El sesgo de género y su influencia en el ámbito laboral", en: *Grupo Ático 34*, Madrid (28.10.2021) https://protecciondatos-lopd.com/empresas/sesgo-de-genero/

— (2021b), "Inteligencia Artificial e igualdad de género", en: *Grupo Ático 34*, Madrid (24.11.2021) https://protecciondatos-lopd.com/empresas/inteligencia-artificial-igualdad-genero/

Ricoy Casas, Rosa María (2021), **"**Sesgos y algoritmos: Inteligencia de género", en: Bonorino Ramírez, Pablo Raúl, Patricia Valcárcel Fernández, Rafael Fernández Acevedo (coords.), *Nuevas normatividades. Inteligencia artificial, derecho y género,* Thomson Reuters Aranzadi, Navarra, Cizur Menor, 89-120.

Rodríguez Canfranc, Pablo (2021), "Galatea o el problema de género en la inteligencia artificial", en: *Telos*, Fundación Telefónica (23.06.2021). https://telos.fundaciontelefonica.com/la-cofa/galatea-o-el-problema-de-genero-en-la-inteligencia-artificial/

Rodríguez Martínez, Marta (2020), "El sexismo de los algoritmos puede hacernos retroceder décadas en igualdad", en: *euronews.com* (10.03.2020) https://es.euronews.com/2020/03/06/como-los-algoritmos-nos-pueden-hacer-retroceder-decadas-en-igualdad-de-genero

Romero, Alexis, Manuel Sánchez (2021), "Orientación, historial laboral único y uso de la inteligencia artificial: el Gobierno aprueba la nueva ley del empleo", en: *Publico.es* (03.12.2021). https://www.publico.es/politica/orientacion-historial-laboral-unico-y.html/amp

Salas, Javier (2017), "Si está en la cocina, es una mujer: cómo los algoritmos refuerzan los prejuicios", en: *Diario El País*, Madrid (22.09.2017). https://elpais.com/elpais/2017/09/19/ciencia/1505818015_847097.html

Smith, Chris *et al.* (2006), "The History of Artificial Intelligence", en: *History of Computing,* CSEP 590ª, 1-27. https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf

Soriano Arnanz, Alba (2021), "La aplicación del marco jurídico europeo en materia de igualdad y no discriminación al uso de aplicaciones de Inteligencia artificial", en: Bonorino Ramírez, Pablo Raúl, Patricia Valcárcel Fernández, Rafael Fernández Acevedo (Coords.), *Nuevas normatividades. Inteligencia artificial, derecho y género,* Thomson Reuters Aranzadi, Navarra, Cizur Menor, 63-88.

Sunstein, Cass R., "Algorithms, Correcting Biases" (December 12, 2018). *Forthcoming, Social Research*, Available at SSRN: https://ssrn.com/abstract=3300171

Tolosa, Paloma, Dibo, Camila (2021), "Inteligencia artificial, discriminación por género y derecho: viejos problemas, nuevos desafíos", en: *Inteligencia Artificial, Tecnologías emergentes y Derecho. Reflexiones interdisciplinarias*, 1, 39-84.

Vantin, Serena (2021), "Inteligencia Artificial y Derecho discriminatorio", en: Llano Alonso, Fernando Higinio, Joaquín Garrido Martín (eds.), *Inteligencia Artificial y Derecho. El jurista ante los retos de la era digital,* Thomson Reuters Aranzadi, Navarra, 367-384.

Venedis, Raphaële, Linda Senden (2020), "EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination", en: Bernitz, Ulf *et al.* (eds.). *General Principles of EU law and the EU Digital Order*, Wolters Kluwer, Alphen aan den Rjn, 151-182.

Verma, Sahil, Julia Rubin (2018). "Fairness Definitions Explained", en: *IEEE/ACM International Workshop on Software Fairness* (FairWare), 1-7. https://fairware.cs.umass.edu/papers/Verma.pdf

Wajcman, Judy (2004), *Technofeminism*, Polity Press, Cambridge.

Warat, Luis Alberto (2000), *Por quem cantam as sereias. Informe sobre Ecocidadania, Gênero e Direito*, Síntese, Porto Alegre.

Yuste, Rafael, Sara Goering, Blaise Agüera y Arcas *et al.* (2017), "Four ethical priorities for neurotechnologies and AI", en: *Nature* 551, 159-163 (09.11.2017) https://doi.org/10.1038/551159a

Informes y Webs

Diagnóstico de la igualdad de género en el medio rural. Ministerio de Medio Ambiente, Medio Rural y Marino, 2011. https://agroinformacion.com/documento-divulgativo-del-%C2%93 diagnstico-de-la-igualdad-de-gnero-en-el-medio-rural/

Comunicación de la comisión al Parlamento europeo, al Consejo europeo, al Consejo, al Comité económico y social europeo y al Comité de las regiones inteligencia artificial para Europa. Bruselas. 25.04.2018. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A 237%3AFIN

Estrategia Nacional de Inteligencia Artificial. ENIA. Ministerio de Asuntos económicos y Transformación Digital. Gobierno de España. Noviembre 2020. https://portal.mineco.gob.es/es-es/ministerio/areas-prioritarias/ Paginas/inteligencia-artificial.aspx

Resolución del Parlamento Europeo, de 14 de marzo de 2017, sobre la igualdad entre mujeres y hombres en la Unión Europea en 2014-2015 (2016/2249(INI)). https://op.europa.eu/en/publication-detail/-/publication/ ed819bab-8fef-11e8-8bc1-01aa75ed71a1/language-es

FRA- Fundamental Rigths Agency of the European Union. Manual de Legislación europea contra la discriminación, 2018.

UNESCO. I'd Blush If I Could. Closin gender divides in digital skills trough education. UNESCO and EQUALS, Skills Coalition. 2020. https://en.unesco.org/Id-blush-if-I-could

UNESCO: Artificial intelligence and gender equality: key findings of Unesco's global dialogue. https://unesdoc.unesco.org/ark:/48223/pf0000374174

"Me sonrojaría si pudiera". El sexismo oculto de los asistentes virtuales (28.04.2020). https://epale.ec.europa.eu/es/content/me-sonrojaria-si-pudiera-el-sexismo-oculto-de-los-asistentes-virtuales

https://www.elespanol.com/espana/20211017/nuevos-jueces-cursos-transexualidad-enjuiciamiento-perspectiva-genero/619688940_0.html

¿Cómo se audita un algoritmo? Los cinco pasos de una auditoría algorítmica. https://www.eticasconsulting.com/como-se-audita-un-algoritmo-pasos-para-auditar-algoritmos/

Noticias ONU. La ausencia de mujeres en el campo de la inteligencia artificial reproduce el sexismo (03.06.2019). https://news.un.org/es/story/2019/06/1456961

Sentencia 13 de julio de 2016: *State v. Loomis*, 881, N.W.2d 749, 7532 (Wis, 2016).

Entrevista de Agnés Bardon. "Hay que educar a los algoritmos", en: *Correo de la UNESCO*. https://es.unesco.org/courier/2020-4/hay-que-educar-algoritmos

How well do IBM, Microsoft, and Face ++ AI Services guess the gender of a face? http://gendershades.org/

Eticas Consulting. Guía de Auditoría Algorítmica. https://www.diariojuridico.com/guia-de-auditoria-algoritmica-para-que-la-ia-cumpla-con-la-legalidad/

Sistema VioGén. Ministerio del Interior. Gobierno de España. http://www.interior.gob.es/web/servicios-al-ciudadano/violencia-contra-la-mujer/sistema-viogen

"Auditando el algoritmo contra la violencia de género", en: *Eticas*. https://eticasfoundation.org/es/eticas-consulting-comienza-una-auditoria-externa-de-viogen-el-algoritmo-del-gobierno-que-asigna-riesgo-a-las-mujeres-victimas-de-violencia-de-genero/

The Montreal Declaration for the Responsible Development of Artificial Intelligence Launched. Canada-Asean Business Council. https://www.canasean.com/the-montreal-declaration-for-the-responsible-development-of-artificial-intelligence-launched/

Recommendation of the Council on Artificial Intelligence. OCDE Legal Instruments. 22.05.2019. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

OECD (2019), *Artificial Intelligence in Society*, OECD Publishing. Paris. https://doi.org/10.1787/eedfee77-en

G20 Ministerial Statement on Trade and Digital Economy. https://www.mofa.go.jp/files/000486596.pdf

Proyecto de texto de la Recomendación sobre la ética de la Inteligencia Artificial. 41 C/73. Informe de la Comisión de Ciencias Sociales y Humanas (SHS). UNESCO. Conferencia General, 41st, 2021 [793] 41ª Reunión. París. 22.11.2021. https://unesdoc.unesco.org/ark:/48223/pf0000379920_spa

Déclaration européenne sur les droits et principes numériques pour la décennie numérique.Comisión Europea. Bruxelles, le 26.1.2022 COM(2022) 28 final https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_fr

Civio, 2 de julio de 2019. https://civio.es/tu-derecho-a-saber/2019/05/16/la-aplicacion-del-bono-social-del-gobierno-niega-la-ayuda-a-personas-que-tienen-derecho-a-ella/

Declaración de Ética IA-LATAM para el diseño, desarrollo y uso de la Inteligencia Artificial https://ia-latam.com/etica-ia-latam/

OHCHR Commissioned Report. Gender stereotyping as a human rights violation. Octubre 2013. https://villaverde.com.ar/nueva-recomendacion-general-nro-33-sobre-acceso-de-las-mujeres-a-la-justicia-cedaw/

REPORT with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies. (2020/2012(INL)). European Parliament. A9-0186/2020. 08.10.2020. https://www.europarl.europa.eu/doceo/document/A-9-2020-0186_EN.html

Propuesta de Reglamento del parlamento europeo y del consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (ley de inteligencia artificial) y se modifican determinados actos legislativos de la unión. bruselas, 21.04.2021. https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0008.02/DOC_1&format=PDF

Carta ética europea sobre el uso de la inteligencia artificial en los sistemas judiciales y su entorno, adoptado por el CEPEJ durante su 31 S t Reunión plenaria. (Estrasburgo, 3-4 de diciembre de 2018). https://campusialab.com.ar/wp-content/uploads/2020/07/Carta-e%CC%81tica-europea-sobre-el-uso-de-la-IA-en-los-sistemas-judiciales-.pdf

IBM Policy Lab. "Inteligencia Artificial: un equilibrio entre regulación y la autorregulación" https://www.ibm.com/blogs/policy/latin-america/wp-content/uploads/sites/6/2020/01/IBM-Policy-Lab-AI-PoV_ES.pdf

CAPÍTULO III

REFLEXIONES SOBRE JUSTICIA, HUMANIDAD Y DIGITALIZACIÓN

STEFANO BINI
Universidad de Córdoba
sbini@uco.es

«Iustitia est constans et perpetua voluntas ius suum cuique tribuens»
(Justiniano)

1. INTRODUCCIÓN

La lentitud, y (por lo tanto) la ineficacia, son bastante a menudo, individuados como los dos perfiles más emblemáticos de la dimensión patológica que afecta -en general y prescindiendo de un específico ordenamiento jurídico- a los sistemas judiciales (Mora-Sanguinetti, 2021, 73).

Frente a estas cuestiones, el proceso de digitalización de la justicia es considerado como una oportunidad extraordinariamente positiva, *«a toolbox of opportunities»* [COM (2020), 710, final], para solucionar problemas estructurales de los sistemas judiciales, ya que «reviste [*rectius*, ha revestido] especial importancia para mantener los órganos judiciales en funcionamiento durante la pandemia de COVID-19 y, más en general, para promover la eficiencia y accesibilidad de los sistemas judiciales» [COM (2021), 389, final, 1].

Pues bien, el presente estudio presenta un razonamiento que se centra en un perfil específico, dentro del macro tema de la digitalización de la sociedad en general y de la justicia en particular: a través de las páginas que siguen, se pretende, de hecho, ofrecer un breve itinerario de investigación alrededor de una cuestión, cuyo relieve parece hoy en día particularmente central y significativa.

¿La tensión entre humanidad y artificialidad, que en cierto modo impregna emblemáticamente la existencia contemporánea, en el ámbito judicial, en torno a qué punto de equilibrio puede "asentarse", para que la fundamental función social expresada por la justicia no se vea desvirtuada por la lógica eficientista alimentada por las tecnologías digitales? ¿Qué punto de *trade-off* puede dibujarse? Y, ¿cómo encontrar este punto de equilibrio -al final- entre humanidad y artificialidad, en la gestión y el funcionamiento del sistema judicial contemporáneo?

La cuestión implica, evidentemente, un necesario razonamiento previo en torno al tema del binomio decisión humano-digital: ¿dónde se sitúa el punto de *trade-off* entre la búsqueda de eficiencia y previsibilidad de las decisiones

(judiciales) digitales, por una parte, y, por otra parte, la intrínseca e imprescindible necesidad de humanidad de las mismas decisiones?

Con respecto a esta cuestión, la tesis que se defiende se basa en la exigencia de destacar la naturaleza necesaria e insuperablemente instrumental de la inteligencia artificial y de las tecnologías digitales en general, para la consecución de objetivos de eficacia y eficiencia de los sistemas judiciales contemporáneos, rechazando claramente visiones que conciban la *"digital transformation"* como fin.

Antes de entrar en el corazón de la reflexión que aquí se propone, parece oportuno desarrollar una premisa conceptual de no despreciable relieve, que tiene por objeto una aclaración o, más correctamente, una opción semántica fundamental.

En el presente estudio se prefiere utilizar la locución "sistema judicial digital" e/o "administración digital de la justicia": a pesar de ser menos evocativa, intuitiva e inmediata de la de "justicia digital", ella parece, efectivamente, expresar mejor el perfil sistemático y administrativo del proceso de digitalización de la misma justicia.

Como es sabido, la idea de "justicia" hunde sus raíces en el terreno de la filosofía (*ex plurimis*, véase Maffettone/Veca, 1997): pensemos en las contribuciones de inestimable valor elaboradas por Platón (con su "República"), Aristóteles (con su "Ética Nicomáquea"), Hobbes (con su "Leviatán"), Rousseau (con su "Discurso sobre el origen de la desigualdad entre los hombres"), hasta llegar a Hayek (con su "Derecho, legislación y libertad") y Rawls (con su "Una teoría de la justicia").

Concepto distinto es el de "administración de la justicia": una locución que se proyecta en la dimensión más concreta y de gestión de la misma justicia, a través de un sistema de normas y principios que dibujan un sistema.

2. EL ESCENARIO DE REFERENCIA. LA CONSTRUCCIÓN DE UNA NUEVA FASE: JUSTICIA DIGITAL Y RECUPERACIÓN POST-PANDÉMICA

La breve reflexión que se desarrolla a continuación quiere ser, en cierto modo, la continuación de otra anterior (Bini, 2021), sobre el impacto que la inteligencia artificial está produciendo en la práctica profesional del Derecho en general.

En aquella reflexión, el razonamiento se desarrollaba a lo largo de dos líneas, indisolublemente unidas entre sí: la reconfiguración de la actividad profesional del abogado, por un lado, y la reestructuración de los procesos de trabajo en los despachos de abogados, por el otro.

Se consideró el cambio digital en el ámbito de la abogacía, prestando especial atención al perfil del paso desde la artesanía a la lógica de los procesos (Susskind, 2017 y 2020) en el marco de una general reestructuración de los procesos de trabajo en el despacho profesional y del contenido de la misma actividad del abogado.

Pues bien, punto de partida de la presente reflexión es -hoy como ayer- una mirada rápida al escenario de referencia.

Un escenario que, en su ser pandémico y post-pandémico, se caracteriza por una marcada orientación hacia la reconstrucción, la recuperación y la reconfiguración: tendencias (*rectius*, necesidades), estas, determinadas, sobre todo, por la crisis económica y social que surgió de la emergencia sanitaria por COVID-19.

Dentro de esta visión general, la digitalización desempeña un papel absolutamente estratégico, de cara a «la recuperación socioeconómica tras la pandemia» (Flechoso, 2021): la «configuración de un futuro digital en Europa» [COM (2020), 67, final] y en el mundo representa, de hecho, una verdadera necesidad ineludible, en consideración de la invasiva y disruptiva difusión de instrumentos y tecnologías digitales en todos los sectores de la vida.

Pues bien, la importancia estratégica del sistema judicial en la consecución de los objetivos de recuperación post-pandémica es evidente, representando uno de los pilares fundamentales de un maduro proceso de modernización de los ordenamientos, que se entrelaza con las trayectorias de la digitalización. ¿Pero qué tipo de modernización, en concreto, se puede (y se debe) conseguir? ¿Qué objetivos, qué estrategias, qué acciones, qué programas?

Al respecto, interesantes perspectivas hermenéuticas pueden encontrarse en el plan de trabajo "Justicia 2030, trasformando el ecosistema del Servicio Público de Justicia": «un plan de trabajo común a 10 años, desarrollado en cogobernanza, que impulsa el Estado de Derecho y el acceso a la Justicia como palancas de la transformación de país» (Ministerio de Justicia, 2021, 3).

La idea de contribuir a la transformación y mejora del país, también a través de una estrategia integrada de intervención en el ámbito de la justicia, con un fuerte anclaje en el «marco conceptual de los Objetivos de Desarrollo Sostenible y de los fondos *Next Generation* EU» (Ministerio de Justicia, 2021, 3) presenta perfiles de interés, consagrando el papel clave de un funcionamiento eficaz del sistema judicial, con vistas a una innovación y una transformación de 360º de la sociedad.

El mismo plan aporta una contribución significativa, entre otras cosas, dibujando los tres objetivos claros y puntuales de una intervención que

-también mediante el uso sistemático de las tecnologías digitales- pretende transformar el Servicio Público de Justicia: 1. mejorar el «acceso al ejercicio de derechos y libertades»; 2. impulsar la «eficiencia del Servicio Público de Justicia»; 3. «contribuir a la sostenibilidad y la cohesión» (Ministerio de Justicia, 2021, 4-6).

Cada objetivo se articula en torno a tres programas, que abarcan un conjunto de proyectos y subproyectos concretos y específicos: así, en lo que se refiere al objetivo número 2, los programas de intervención se sitúan en tres dimensiones de la eficacia de la justicia: organizativa, procesal y digital.

En otras palabras, la eficiencia y la modernización del sistema judicial de un ordenamiento jurídico, como claves para la transformación y la innovación de un país, se fundan en tres pilares estratégicos, que identifican los fundamentales ámbitos de intervención: la organización, el proceso y el digital.

3. MODERNIZAR LOS SISTEMAS JUDICIALES DE LA UNIÓN EUROPEA

Precisamente en materia de "modernización de los sistemas judiciales", parece de extraordinario interés mirar al escenario europeo, identificando dos fechas como emblemáticas: el 2 de diciembre de 2020 y el 1 de diciembre de 2021.

En esta última, la Comisión Europea ha adoptado un conjunto de iniciativas orientadas a la digitalización de los diferentes sistemas judiciales de la Unión, para que sean más accesibles y eficaces, a través de la digitalización de la cooperación judicial transfronteriza, del intercambio digital de información en casos de terrorismo y de la creación de una plataforma de colaboración para los equipos conjuntos de investigación.

«El objetivo general de las medidas es convertir los canales de comunicación digital en el canal por defecto en los asuntos judiciales transfronterizos, llevando así a la práctica una de las prioridades establecidas el año pasado en la Comunicación sobre la digitalización de la justicia» (Comisión Europea, 1 de diciembre de 2021).

Y, efectivamente, la intervención de 2021 se pone en plena coherencia con la adopción, el año anterior, de un conjunto de medidas animadas por la finalidad de impulsar la digitalización de los sistemas judiciales, tanto a nivel nacional como comunitario. Más en detalle, procede destacar la importancia de la Comunicación de la Comisión sobre la digitalización de la justicia (*"Digitalisation of justice in the European Union A toolbox of opportunities"*: "un abanico de oportunidades") y de la Comunicación de la Comisión sobre "Garantizar la justicia en la UE: estrategia europea sobre la formación judicial

para 2021-2024", ambas de 2020 [COM (2020), 710, final y COM (2020), 713, final].

Pues bien, de estas dos intervenciones de la Comisión Europea se desprende claramente la visión europea en materia de digitalización de los sistemas judiciales: una visión que se basa en el carácter instrumental de las tecnologías y en la facilitación del acceso y de la puesta en común de las informaciones (los datos, los *big data* que alimentan el funcionamiento de las inteligencias artificiales) [COM (2020), 710, final].

Y una visión que identifica asimismo en la formación judicial [COM (2020), 713, final] un pilar fundamental, verdaderamente imprescindible para una efectiva modernización de los sistemas judiciales (al respecto, muchos son los puntos de contacto con el "Plan de Educación Digital Europeo", cuyo desarrollo representa una de las grandes urgencias de la Unión Europea: cf. Gómez Muñoz, 2021, 90).

4. RIESGOS ARTIFICIALES Y GARANTIAS HUMANAS

Como pone de manifiesto la Comunicación de la Comisión "Digitalisation of justice in the European Union. A toolbox of opportunities" [COM (2020), 710, final], «While the advantages of introducing AI-based applications in the justice system are clear, there are also considerable risks associated with their use for automated decision-making and 'predictive policing' / 'predictive justice'». Es decir, «si bien las ventajas de introducir aplicaciones basadas en la inteligencia artificial en el sistema de justicia son claras, también hay riesgos considerables asociados con su uso para la toma de decisiones automatizada y la prevención policial/ justicia predictiva» [COM (2020), 710, final, 11].

Y precisamente la noción de "riesgo" resulta central en la construcción de una "Europa adaptada a la era digital", según el planteamiento presentado por la Comisión Europea en la "Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de inteligencia artificial) y se modifican determinados actos legislativos de la Unión", del abril de 2022 [COM (2021), 206, final].

Se trata de una contribución particularmente significativa, que aspira a «convertir a Europa en el centro mundial de una inteligencia artificial (IA) digna de confianza» (Comisión Europea, comunicado de prensa, 21 de abril de 2021). Para que este objetivo pueda ser en concreto alcanzado, se elabora un planteamiento basado en el concepto de "riesgo", clasificado en el marco de una taxonomía que diferencia y ordena los diferentes sistemas de inteligencia artificial según el nivel de riesgo que llevan consigo (*unacceptable, high, limited, minimal*).

Sin tomar en consideración todas las diferentes categorías de riesgo, basta con destacar que, si es "inadmisible" (*"unacceptable"*) el riesgo que conllevan los sistemas de inteligencia artificial que representen una amenaza para los derechos y la seguridad de la persona, incluso llegando a concretar formas de manipulación del comportamiento humano, se considera, en cambio, "alto" el riesgo que caracteriza los sistemas de inteligencia artificial utilizados en sectores neurálgicos como -por ejemplo- los de las infraestructuras críticas (pensemos en los transportes), de la formación educativa y profesional, del empleo, de la gestión de las personas que trabajan o de la migración y de la administración de justicia y procesos democráticos.

Con respecto a estos sistemas de inteligencia artificial, el planteamiento comunitario en materia prevé la sujeción a obligaciones "estrictas", que se basan en conceptos básicos de fundamental importancia, entre los cuales cabe señalar: la solidez, la seguridad, la precisión, la claridad en las dinámicas de funcionamiento de los mismos sistemas; la adecuación de la información proporcionada a los usuarios (con la relativa documentación); la trazabilidad de los resultados y de la actividad; la evaluación y la mitigación de los riesgos. Un perfil específico merece ser puesto en luz con especial atención: la propuesta de regulación, siempre con referencia a los sistemas de IA de alto riesgo, prevé también que estos sean sujetos a la obligación de adoptar «medidas apropiadas de supervisión humana para minimizar el riesgo» (Comisión Europea, 21 de abril de 2021).

Este aspecto reviste especial relevancia en consideración de la visión estratégica que expresa y que puede, esencialmente, sintetizarse en la centralidad del carácter instrumental y, por tanto, en el enfoque antropocéntrico de la tecnología digital. Efectivamente, en la misma previsión de la supervisión humana sobre el funcionamiento de los sistemas de inteligencia artificial de riesgo alto, se puede decir consagrado un planteamiento basado en la idea de "límite".

Idea de límite que encuentra en la norma jurídica en sentido amplio su natural proyección instrumental, para la consecución de una efectiva protección de los derechos de la persona. Como recuerda magistralmente Alain Supiot: «la funzione primordiale del Diritto è proprio quella di tracciare dei limiti che siano al contempo sufficientemente robusti per poter contenere la legge del più forte, e sufficientemente chiari e precisi per consentire agli uomini di esercitare, su un piano di parità, le loro libertà individuali e collettive» (Supiot, 2020, 155).

Esta necesidad se entiende de manera especial, sobre todo si se toma en consideración el escenario específico de la administración de la justicia. De hecho, muchos son los perfiles de riesgo y los aspectos críticos que surgen en el estudio de la aplicación de las tecnologías basadas en inteligencias artificiales

de ultimísima generación (los así llamados *"machine learning"* y *"deep learning"*) a los procesos de toma de decisiones en el ámbito judicial. Merecen ser estudiados atentamente y neutralizados a través de un aparato normativo que, más allá de los perímetros nacionales, funcione como "dosificador": «ante los riesgos de la economía digital el papel de la regulación jurídica (…) es necesario como "dosificador" o "limitador" de velocidad" de los cambios, para evitar la desconexión entre las tecnologías, las personas y las instituciones (…)» (Navarro Nieto, 2021, 25).

Pues bien, con respecto al específico ámbito de la presente investigación, perfiles de posible desconexión entre tecnología digital y dimensión humana pueden ser detectadas en algunas experiencias de los así llamados "tribunales de internet".

En 2017/2018 fue bastante llamativa y disruptiva la noticia de la institución, en el ordenamiento jurídico chino, de los así llamados "tribunales de internet": órganos jurisdiccionales competentes, exclusiva e íntegramente, en litigios relacionados con internet en sentido amplio. Pensemos en la así llamada "Hangzhou Internet Court": un tribunal *internet-related* (v. https://youtu.be/QkczNbGxvN4; fecha de última consulta: 8 de junio de 2022).

Estos peculiares e innovadores "espacios" (*rectius*, órganos) judiciales se caracterizan por ser relacionados con la tecnología digital, desde un punto de vista no sólo de competencia, sino también metodológico-instrumental, siendo, de hecho, su funcionamiento centrado en la plataforma digital y, por lo tanto, en el algoritmo que, en algunos casos, puede llegar incluso a juzgar.

Más en general, «l'intelligenza artificiale sta arrivando nel mondo degli avvocati e in quello dei tribunali, con alcuni giudici che negli Stati Uniti cominciano a usare algoritmi per stabilire la pena da infliggere ai condannati» (Gaggi, 2018, 46).

5. CONCLUSIONES

Los algoritmos se presentan, así, como posible solución al problema de la incertidumbre y, por lo tanto, como instrumento para detener el perímetro de la duda en la justicia. Precisamente al respecto, parece oportuno preguntarse: ¿hay todavía espacio para las dudas? Como escribe Gianrico Carofiglio, «dudar es un arte práctico. Y como todo arte práctico, se aprende. Dudar, hacer y hacerse preguntas es el único medio para llegar a conocer lo que se desconoce (por ejemplo -pero naturalmente no sólo- la verdad procesal)» (Carofiglio, 2010, contraportada).

Pues bien, como ha sido muy oportunamente destacado en el horizonte doctrinal francés, «la giustizia digitale è il teatro di uno scontro fra due modi di

produrre senso e di organizzare la coesistenza umana, che abbiamo chiamato (…) due "forme simboliche": il diritto e il digitale» (Garapon/Lassègue, 2021, 276).

Entre estos dos términos conceptuales, puede, de alguna manera, reconocerse la existencia de una natural tensión, que se declina -entre otras cosas- en el binomio: certidumbre artificial/incertidumbre humana.

A una evaluación crítica no puede escapar que, si bien es cierto que la celeridad del desarrollo procesal representa un elemento fundamental, condicionante de la misma efectividad de la tutela judicial, también es cierto que «il diritto è un discorso, un'arte di vivere insieme alla giusta distanza, di distribuire status, beni, riconoscimenti, necessariamente debitrice verso un'antropologia che a monte plasma costumi, usanze e credenze, un'economia che produce e accumula le ricchezze e un potere che le garantisce» (Garapon/Lassègue, 2021, 276; cf. Carnelutti, 2017).

Así que reducir la complejidad de la cuestión objeto de estudio a la simple oportunidad de privilegiar la certidumbre y la rapidez garantizadas por la dimensión científica del algoritmo representa, francamente, una opción equivocada y no coherente con una visión sistemática del Derecho. De hecho, parece, más bien, oportuno y necesario, abrazar un enfoque que sitúe la cuestión dentro de un contexto ordinamental intrínseca e insuperablemente humano-céntrico.

La tecnología digital constituye una extraordinaria oportunidad, tanto en general, como en particular en el ámbito de la administración de la justicia, pero dentro de ciertos límites: en otras palabras, se considera indispensable mantener la dimensión instrumental de la tecnología, sin que pueda convertirse -como, en realidad, ocurre en algunos modelos *infra* mencionados- el instrumento en sujeto del proceso decisional judicial.

Como escribe Natalino Irti, «il diritto è il mondo della decisione» y «la decisione è sempre una scelta, un atto selettivo. (…) Contraddittorio, dubbio e decisione costituiscono fasi di un processo unitario (…)» (Irti, 2016, 117 y 119). Pues bien, precisamente el tema de la toma de decisiones representa probablemente el punto neurálgico de una cuestión que parece trascendental para la construcción de una sociedad digital realmente sostenible y que siga teniendo su enfoque en la persona.

Al respecto, se comparte plenamente el planteamiento que la Comisión Europea expresa en la Comunicación sobre la digitalización de la justicia: «the final decision-making must remain a human-driven activity and decision. Only a judge can guarantee genuine respect for fundamental rights, balance conflicting interests and reflect the constant changes in society in the analysis of a case. At the same time, it is important that judgments are delivered by judges

who fully understand the AI applications and all information taken into account therein that they might use in their work, so that they can explain their decisions» [COM (2020), 710, final].

La toma de decisiones -en general y de manera especial en el ámbito judicial- no puede que ser intrínseca e insuperablemente humana, ya que «sólo un juez puede garantizar un verdadero respeto de los derechos fundamentales, equilibrar los intereses en conflicto y reflejar los cambios constantes en la sociedad en el análisis de un caso. Al mismo tiempo, es importante que las sentencias sean dictadas por jueces que comprendan plenamente las solicitudes de AI y toda la información que en ellas se tenga en cuenta que puedan utilizar en su trabajo, para que puedan explicar sus decisiones» (traducción).

En definitiva, «la transformación digital de la justicia es un proceso que ya se ha iniciado, pero en el que todavía queda mucho camino por recorrer» (Delgado Martín, 2022, 1); un camino que exige que se tenga muy en cuenta -en una dimensión de «derecho global» (Cassese, 2009)- el alcance de los «desafíos que la tecnología ofrece a la justicia: tanto en la implantación de soluciones tecnológicas para mejorar el sistema judicial, especialmente aquellas que tienen componentes disruptivos (inteligencia artificial, plataformas *online* de resolución de litigios y tecnología *blockchain*); como en la protección frente a los ataques más graves procedentes de las tecnologías (…)» (Delgado Martín, 2022, 1).

En un ámbito más -el de la justicia y del funcionamiento de sus sistemas- se advierte la inaplazable necesidad de poner límites a la digitalización descontrolada y a la búsqueda desenfrenada de la sola eficiencia en el funcionamiento de los sistemas judiciales, coherente con una visión de la administración de la justicia que concibe la duda como un obstáculo a la misma eficiencia, olvidando que «il problema del dubbio non è un frívolo divertimento di giuristi; esso nasce nel seno stesso del diritto positivo» (Irti, 2016, 123; cf. Giabardo, 2020).

6. BIBLIOGRAFÍA

Bini, Stefano (2021), "Algoritmos y abogacía digital: reflexiones sobre el cambio de paradigma en el trabajo del abogado contemporáneo", en: Llano Alonso, Fernando, Joaquín Garrido Martín (eds.), *Inteligencia artificial y Derecho. El jurista ante los retos de la era digital*, Cizur Menor, Aranzadi, 51-65.

Carnelutti, Francesco (2017), *Arte del diritto*, Giappichelli, Torino.

Carofiglio, Gianrico (2010), *El arte de la duda*, Marcial Pons, Madrid.

Cassese, Sabino (2009), *Il diritto globale. Giustizia e democrazia oltre lo Stato*, Einaudi, Torino.

Delgado Martín, Joaquín (2022), "La transformación digital de la justicia es un proceso que ya se ha iniciado, pero en el que todavía queda mucho camino por recorrer", en: *Diario La Ley*, 17 de febrero.

Flechoso, José Joaquín (2021), *Digitalización y recuperación económica. El papel de la digitalización en la recuperación socioeconómica tras la pandemia*, Almuzara, Córdoba.

Gaggi, Massimo (2018), *Homo Premium. Come la tecnologia ci divide*, Laterza, Bari.

Garapon Antoine, Jean Lassègue (2021), *La giustizia digitale*, Il Mulino, Bologna.

Giabardo Carlo Vittorio (2020), *Il giudice e l'algoritmo (in difesa dell'umanità del giudicare)*, en: *Giustizia insieme*, www.giustiziainsieme.it (1/5/2021).

Gómez Muñoz José Manuel (2021), "La acción de la Unión Europea durante la pandemia de SARS-COV-2", en: José Manuel Gómez Muñoz (ed.), *Nuevos escenarios del sistema de relaciones laborales derivados del COVID-19*, Bomarzo, Albacete, 65-91.

Irti, Natalino (2016), *Un diritto incalcolabile*, Giappichelli, Torino.

Maffettone, Sebastiano, Veca Salvatore (eds.) (1997), *L'idea di giustizia da Platone a Rawls*, Laterza, Bari.

Mora-Sanguinetti, Juan S (2021), *La factura de la injusticia*, Tecnos.

Navarro Nieto, Federico (2021), "Representación y acción sindical en la economía digital", en: José Manuel Gómez Muñoz (dir.), *Sindicalismo y capitalismo digital: los límites del conflicto*, Bomarzo, Albacete, 15-61.

Susskind, Richard (2017), *El abogado del mañana. Una introducción a tu futuro*, Wolters Kluwer, Madrid.

— (2020), *Tribunales online y la Justicia del futuro*, Wolters Kluwer, Madrid.

Fuentes, informes y documentos de interés citados en el texto

Comisión Europea (2020), *Comunicación al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones "Configurar el futuro digital de Europa"* [COM(2020) 67 final].

Comisión Europea (2020), *Comunicación al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones "La digitalización de la justicia en la UE: Un abanico de oportunidades"* [COM(2020) 710 final].

Comisión Europea (2020), *Comunicación al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones "Garantizar la justicia en la UE: estrategia europea sobre la formación judicial para 2021-2024"* [COM(2020) 713 final].

Comisión Europea (2021), *Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de inteligencia artificial) y se modifican determinados actos legislativos de la Unión"* [COM(2021) 206 final].

Comisión Europea (2021), *Comunicación al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones "Cuadro de indicadores de la justicia en la UE de 2021"*, [COM(2021) 389 final].

Comisión Europea (2021), *Modernizar la cooperación judicial: la Comisión prepara el camino para una mayor digitalización de los sistemas judiciales de la UE*, comunicado de prensa, 1 de diciembre.

Comisión Europea (2021), *Una Europa adaptada a la Era Digital: la Comisión propone nuevas normas y medidas para favorecer la excelencia y la confianza en la inteligencia artificial*, comunicado de prensa, 21 de abril.

Gobierno de España, Ministerio de Justicia (2021), *Justicia 2030. Trasformando el ecosistema del Servicio Público de Justicia. Resumen ejecutivo*.

Contribución audiovisual *An exclusive tour of China's first Internet Court in Hangzhou*, https://youtu.be/QkczNbGxvN4 (fecha de última consulta: 8 de junio de 2022).

CAPÍTULO IV

INTELIGENCIA ARTIFICIAL Y DERECHOS FUNDAMENTALES

LAURA GÓMEZ ABEJA
Universidad de Sevilla
lgomez2@us.es

1. INTRODUCCIÓN

Nuestras vidas han cambiado mucho en los últimos años gracias a los avances tecnológicos, que se encuentran estrechamente vinculados al uso del *big data* -también conocido como "macrodatos" o "datos masivos"- y a la inteligencia artificial. En efecto, la ingente -y siempre creciente- cantidad de datos disponible y los avances informáticos y algorítmicos han permitido a la inteligencia artificial facilitar nuestro día a día, con aplicaciones que nos ayudan -por ejemplo- a llegar a un lugar determinado, a saber cuántas calorías hemos consumido, o a identificar una canción que oímos y nos agrada, pero cuyo título y autor no conocemos o no recordamos; pero la IA también ha servido para afrontar grandes problemas a los que nos enfrentamos hoy como sociedad, como el tratamiento de algunas enfermedades crónicas, la prevención de otras graves, o la lucha contra el cambio climático[1]. Es evidente que desde esta perspectiva debe hacerse una valoración muy positiva de la IA y del *big data*, que han permitido unos avances inimaginables hasta hace pocos años. Pero el uso de estas herramientas también ha traído consigo un aspecto negativo. Y no es un problema menor, pues se trata de la vulneración -también masiva- de ciertos derechos fundamentales que, además, conforme avanza la acumulación y tratamiento de datos y el desarrollo de las nuevas tecnologías mediante la IA, se antoja cada vez más difícil de atajar. En estas páginas se van a hacer unas consideraciones generales al hilo del impacto del *big data* y el uso de algoritmos (para la toma de decisiones que afectan a personas) en dos derechos fundamentales: el derecho a la protección de datos personales (art. 18.4 CE) y el derecho a la no discriminación (art. 14 CE). Para ello, en primer lugar se harán las necesarias precisiones terminológicas. Después se expondrá cómo el uso de esas herramientas afecta en particular a los derechos mencionados. En tercer lugar se incidirá en las medidas -jurídicas- que se han adoptado para garantizar la protección de estos derechos y en algunas de las limitaciones que -a mi juicio- presentan las mismas. Finalmente se efectuarán a modo de conclusión algunos apuntes, con carácter esencialmente propositivo.

[1] Como indica la Comunicación de la Comisión de 25 de abril de 2018, al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo, y al Comité de las Regiones Inteligencia Artificial para Europa [COM (2018) 237 final].

2. PRECISIONES TERMINOLÓGICAS

Muchos de los conceptos relevantes en este contexto son relativamente recientes, o al menos novedosos para la mayoría de nosotros. Esto, junto al hecho de que habitualmente se trate de anglicismos (el de las TICs es uno de los muchos ámbitos en que los conceptos nuevos hacen fortuna en su versión inglesa), nos sitúa ante un sinfín de términos desconocidos -cada vez menos, eso sí- y exóticos, tales como *big data, data mining*, algoritmos, inteligencia artificial, *machine learning*, *privacy by design/by default*, o perfilado, entre muchos otros. Se hace necesario, pues, precisar algunos de estos conceptos.

En primer lugar, con el término *big data* se hace referencia, siguiendo lo dispuesto por el Parlamento Europeo, a "la recopilación, análisis y acumulación constante de grandes cantidades de datos, incluidos datos personales (…), objeto de un tratamiento automatizado mediante algoritmos informáticos y avanzadas técnicas de tratamiento de datos, utilizando tanto datos almacenados como datos transmitidos en flujo continuo, con el fin de generar correlaciones, tendencias y patrones"[2], a lo que debe añadirse que todo ello puede hacerse "a una gran velocidad de tratamiento" (Eguíluz Castañeira, 2020, 328). El concepto *big data* -de lo que también se habla como datos masivos o macrodatos- está asociado, por tanto, al de algoritmo. Con *algoritmo*, en el contexto que nos ocupa y en términos sencillos, nos estaremos refiriendo a la fórmula informatizada que se utiliza para calcular una predicción, mediante el uso de los datos disponibles. Es importante añadir que el algoritmo es una "creación humana" (Eguíluz Castañeira, 2020, 330), en el sentido de que son personas las que elaboran la fórmula que después será ejecutada por una computadora. Otra noción esencial es la de *inteligencia artificial,* con la que, según el Grupo de Expertos en Inteligencia Artificial de la Comisión Europea, se hace referencia a los sistemas que muestran un comportamiento inteligente, analizando su entorno y realizando acciones con cierta autonomía para lograr sus objetivos específicos[3]. La inteligencia artificial no puede entenderse sin la noción de algoritmo, a través del que se articula. Consecuentemente, su virtualidad -como

[2] Resolución del Parlamento Europeo, de 14 de marzo de 2017, sobre las implicaciones de los macrodatos en los derechos fundamentales: privacidad, protección de datos, no discriminación, seguridad y aplicación de la ley (2016/2225(INI)), en línea: https://www.europarl.europa.eu/doceo/document/TA-8-2017-0076_ES.pdf (última consulta: 22 de febrero de 2022).

[3] El Grupo de Expertos de Alto Nivel en Inteligencia Artificial se constituyó siguiendo lo previsto en la Comunicación de la Comisión sobre Inteligencia Artificial para Europa, de 25 de abril de 2018, dirigida al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y social Europeo, y al Comité de las Regiones. El documento, "Una definición de la Inteligencia Artificial: principales capacidades y disciplinas científicas", publicado el 18 de diciembre de 2018, está disponible en línea: https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf (última consulta: 22 de noviembre de 2021).

la de aquél- depende también de la intervención humana, por eso la autonomía de la IA a la que se ha hecho referencia es sólo relativa, en el sentido de que se desarrollará, en principio, dentro de los parámetros previstos por las personas programadoras del sistema y creadoras de los algoritmos que lo integren.

Desde la perspectiva del debate jurídico, no en el sentido de que sean sinónimos desde un punto de vista técnico, por supuesto, algoritmo e inteligencia artificial se han llegado a utilizar indistintamente (Eguíluz Castañeira, 2020, 331). De otra parte, como se ha afirmado, "*big data* e inteligencia artificial se relacionan de manera bidireccional: por un lado, la IA necesita una gran cantidad de datos para el aprendizaje automático y, por el otro, el *big data* utiliza la IA y sus técnicas para extraer valor de los grandes volúmenes de información" (Arellano Toledo, 2019, 5). Los algoritmos y el *big data,* en fin, forman parte del universo de la inteligencia artificial. Todos ellos son, pues, conceptos íntimamente imbricados. A lo largo de estas páginas saldrán a colación otros conceptos propios del tema que nos ocupa. Por el momento basta con conocer los expuestos en las líneas previas.

3. DERECHOS FUNDAMENTALES EN JUEGO

El presente trabajo se centra en dos derechos -la protección de datos y la no discriminación- que resultan especialmente relevantes en este contexto, pero antes de profundizar en ellos debe al menos recordarse que existen otros derechos incididos a consecuencia del uso del *big data* y del desarrollo de la inteligencia artificial y sus avanzados algoritmos para la toma de decisiones, como el derecho a la tutela judicial (Castellanos Claramunt/Montero Caro, 2020), el derecho a la libertad de información (Cotino Hueso, 2017, 140), el derecho al sufragio (García Mahamuth, 2015) o el derecho de acceso a la información pública (Medina Guerrero, 2021).

3.1. El derecho a la protección de datos personales

Uno de los derechos que se ha visto gravemente cuestionado a consecuencia de la acumulación masiva de datos es el derecho a la protección de los datos personales. Gran parte de la información del *big data* son datos de carácter personal. La tecnología permite que empresas privadas y poderes públicos utilicen masivamente datos personales en la realización de sus actividades. Por su parte, las personas físicas divulgan cada vez más información personal al público en general[4]. En este último caso, los datos

[4] Así lo señala el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos), en su considerando sexto.

proceden esencialmente de las redes sociales, pero las fuentes que proporcionan esta clase de información al *big data* son muy diversas, desde los dispositivos inteligentes -en general, el "internet de las cosas"- hasta las empresas con las que se efectúan transacciones, que proporcionarán información sobre pagos o datos administrativos[5]. Como es obvio, las posibilidades de control de esa información por parte de sus titulares se difuminan en un universo de trasiego constante y masivo de datos, por los que todos pujan. Y es que ello sucede sobre todo por el gran interés que existe en los datos, de los que se ha hablado como el petróleo del siglo XXI, pues se han convertido en un activo patrimonial de un valor económico sin precedente en los mercados (Sancho López, 2019, 3):

> Los datos son hoy el propulsor de crecimiento y transformación, como lo fue el petróleo en su momento. Y los flujos de datos configuran hoy nuevas infraestructuras, nuevos modelos de negocio y nuevas economías, con nuevos actores en posición de monopolio y políticas estatales diferenciadas según las ventajas de partida para beneficiarse de las reglas de mercado (Moreno Muñoz, 2017, 9, citado en Sancho López, 2019, 3).

Que el sector de los macrodatos esté "creciendo a un ritmo del cuarenta por ciento anual, siete veces más rápidamente que el del mercado de las tecnologías de la información" (Parlamento Europeo, 2017, considerando K) o que se calculase que en Europa el *big data* implicaría un incremento adicional del PIB de un uno coma nueve por ciento para 2020 (Cotino Hueso, 2019, 9), son elementos que corroboran sin atisbo de duda el valor que han alcanzado los macrodatos. Se ha hablado, en este sentido, del *big data* como "*big deal*: el *big data* es un gran negocio. Los datos personales se capitalizan y se comercializan por completo" (Han, 2014, 98, citado en Sancho López, 2019, 4).

Más allá del valor que tienen actualmente *per se* en los mercados, ¿cómo y para qué se utilizan los datos? Pues con ellos, como explica L. Cotino, se hace una "estadística del todo", combinándose miles de datos de forma prácticamente aleatoria: "frente a la contrastación de una hipótesis a partir de los datos, se descubren correlaciones sin conocer previamente la causa. Así sucede al probar casi aleatoriamente la posible correlación entre datos en principio totalmente distantes" (Cotino Hueso, 2017, 133, citando a Martínez, 2014, 3). Las conexiones resultantes permiten generar patrones de tendencias de futuro, a lo que se puede dar una diversidad de usos, válidos para todos los

[5] Así se ha puesto de manifiesto en el documento de la *European Union Agency for Fundamental Rights*, "*BigData: Discrimination in data-supported decision making*", disponible en línea: https://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-focus-big-data_en.pdf (última consulta: 21 de febrero de 2022).

sectores de actividad (tanto para el ámbito público como privado): mejor conocimiento del cliente, del mercado, personalización de productos y servicios, mejora y rapidez en la toma de decisiones, o previsión del comportamiento (AEPD-ISMS, 2017, 6-7). Y el beneficio que se obtiene por el uso de datos es incontestable: las empresas que basan la toma de decisiones en los conocimientos procedentes de datos experimentan un aumento aproximado de la productividad de un cinco con seis por ciento[6].

Expuesto cuanto antecede, puede comprenderse por qué los datos (que a menudo serán datos de carácter personal) tienen tanto valor y por qué en la actualidad hemos llegado a una suerte de "expropiación de la privacidad sin precedentes" (Sancho López, 2019, 24).

3.2. El derecho a la no discriminación

Otra consecuencia del uso del *big data* y los algoritmos es la posible lesión del derecho a no ser discriminado. Concretamente, en términos del Parlamento Europeo, los macrodatos pueden resultar en un tratamiento diferenciado injustificado "y en una discriminación indirecta de grupos de personas con características similares, en particular en lo que se refiere a la justicia e igualdad de oportunidades en relación con el acceso a la educación y al empleo, al contratar o evaluar a las personas o al determinar los nuevos hábitos de consumo de los usuarios de los medios sociales" (Parlamento Europeo, 2017, consideración general decimonovena). La discriminación se puede producir en distintas fases del proceso de toma de decisiones mediante el uso de datos: desde el momento en que estos son recogidos hasta el de la aplicación del algoritmo por el que se toma la decisión de que se trate.

En un primer momento podría pensarse que estas técnicas presentan una objetividad intachable, pero no es así, o al menos no siempre lo es. Son personas las que fijan las reglas sobre qué datos se van a utilizar y cómo, por lo que estas decisiones inevitablemente tienen una vertiente subjetiva. Las creencias o los valores de quienes deciden qué datos recabar y cómo utilizarlos podrían quedar reflejados en su recogida y tratamiento. En muchas ocasiones el sesgo no será evidente, e incluso puede ocurrir que la persona que trabaja con el sistema en cuestión no esté reflejando en el proceso ideas o valores discriminatorios, pero que el resultado finalmente sí lo sea. Siguiendo el ejemplo de Zuiderveen Borgesius (2018:17), una empresa podría utilizar un algoritmo para la contratación de su personal en el que la "variable objetivo" (que define lo que se quiere encontrar) sea "ser un buen profesional", y usar para su determinación

[6] Así puede leerse en la edición de *El país* de 13 de octubre de 2014, en línea: https://elpais.com/tecnologia/2014/10/13/actualidad/1413199836_461461.html (última consulta: 22 de febrero de 2022).

una "etiqueta de clase" (que divide los datos en categorías) relativa a la puntualidad, en la que se indique que "rara vez llega tarde". Las personas más pobres suelen vivir en el extrarradio y deben viajar más lejos que otros trabajadores para llegar al centro de trabajo. En consecuencia, la gente más pobre suele llegar tarde al trabajo más habitualmente que otros, por los atascos y problemas con el transporte público. Además, estos grupos humanos están integrados muy a menudo por inmigrantes. Si la empresa utiliza la etiqueta "raramente llega tarde" para determinar si un empleado es un buen profesional, se estaría poniendo en desventaja a la población inmigrante y más pobre, aunque sean mejores en otros aspectos (contemplados o no en otras etiquetas).

Puede suceder también, claro está, que el sesgo se introduzca de forma deliberada. Siguiendo con el ejemplo de Zuiderveen (2018: 22), una empresa podría servirse de un algoritmo que permite predecir el embarazo (que utilizan algunas compañías con fines de marketing) para evitar contratar a mujeres embarazadas, discriminándolas deliberadamente[7]. O podrían introducirse sesgos para dirigir determinados comportamientos y -o- crear opiniones políticas, como sucedió en las elecciones estadounidenses con *Cambridge Analytica* (Arellano Toledo, 50, 2019, 8).

El hecho de que quienes determinen la forma en que se van a recabar los datos y quienes creen los algoritmos sean personas supone además que pueda distinguirse un grupo privilegiado en este sentido, el de quienes pueden leer los datos, que pueden impedir a los demás el acceso a esos datos o el acceso al conocimiento generado gracias a los mismos. En el extremo opuesto, siendo objeto de especial discriminación, se encontrarían los marginados del *big data*, que sencillamente no aportan datos al mismo, millones de personas cuyas "preferencias y necesidades están en riesgo de ser ignoradas en las decisiones que se basen en el *big data* y la inteligencia artificial" (Cotino Hueso, 2017, 138).

Puede compartirse pues la afirmación de que "los algoritmos no son imparciales" (Arellano Toledo, 2019, 8). Y debe añadirse que tampoco son exactos. Las personas que trabajan creando los sistemas de inteligencia artificial pueden equivocarse y los errores se reflejarán en los resultados que el sistema arroje; pero es que, además, en contra de la creencia general que atribuye una alta precisión a los sistemas de IA, el *big data* no es siempre fiable. Las interrupciones y las pérdidas de datos que se producen con frecuencia en el mundo virtual producen errores en los resultados algorítmicos (Cotino Hueso, 2017, 134).

[7] Explica el autor (Zuiderveen Borgesius, 2018, 22 y 23) que la cadena estadounidense de tiendas *Target* creó un sistema de predicción de embarazo por puntuación basado en unos veinticinco productos, analizando el comportamiento de los clientes en sus compras. Si una mujer compra algunos de esos productos, *Target* puede predecir con bastante certeza que está embarazada.

Conviene citar -para concluir- las palabras del Parlamento Europeo en su Resolución de 2017:

> Los datos y/o los procedimientos de baja calidad en los que se basan los procesos de toma de decisiones y las herramientas analíticas podrían dar lugar a algoritmos sesgados, correlaciones falsas, errores, una subestimación de las repercusiones éticas, sociales y legales, el riesgo de utilización de los datos con fines discriminatorios o fraudulentos y la marginación del papel de los seres humanos en esos procesos, lo que puede traducirse en procedimientos deficientes de toma de decisiones con repercusiones negativas en las vidas y oportunidades de los ciudadanos, en particular los grupos marginalizados, así como generar un impacto negativo en las sociedades y empresas (Parlamento Europeo, 2017, considerando M).

4. RECONOCIMIENTO DE LOS DERECHOS INCIDIDOS Y MEDIDAS ADOPTADAS RECIENTEMENTE PARA SU PROTECCIÓN

Los derechos a la protección de datos personales y a la no discriminación, como es bien conocido, encuentran amparo en el ordenamiento jurídico español al más alto nivel normativo, en la Constitución, y su contenido ha sido desarrollado y delimitado por el máximo intérprete de la norma fundamental, el Tribunal Constitucional. Algo similar sucede a nivel regional europeo, con las normas *cuasi* constitucionales y los tribunales a los que corresponde su interpretación. El uso de la IA, también es sabido, ha provocado nuevas formas de lesión e invasión en estos derechos. Los mecanismos para su protección han ido quedando obsoletos, y han tenido que crearse nuevas fórmulas para luchar contra la invasión en estas libertades, mediante la aprobación de normas y la creación de órganos *ad hoc*, como se verá enseguida. Especialmente en la Unión Europea se ha llevado a cabo una constante actividad en este sentido desde sus distintas instancias.

4.1. El derecho a la protección de datos personales

4.1.1. Reconocimiento multinivel del derecho

El apartado cuarto del artículo 18 de la Constitución Española establece una genérica limitación legal del "uso de la informática para garantizar el honor y la intimidad personal y familiar de los ciudadanos y el pleno ejercicio de sus derechos", y el Tribunal Constitucional determinó que el mismo ampara un derecho fundamental a la protección de los datos personales; derecho que tiene, por un lado, carácter *instrumental*, al configurarse como un "un instituto de garantía de los derechos a la intimidad y al honor y del pleno disfrute de los restantes derechos de los ciudadanos" (STC 292/2000 FJ 5, que cita la STC 254/1993, FJ 6) y, por otro lado, naturaleza de derecho *autónomo*, que faculta a

"controlar el flujo de informaciones que conciernen a cada persona" (STC 94/1998, FJ 6). El contenido de este derecho "consiste en un poder de disposición y de control sobre los datos personales que faculta a la persona para decidir cuáles de esos datos proporcionar a un tercero, sea el Estado o un particular, o cuáles puede este tercero recabar, y que también permite al individuo saber quién posee esos datos personales y para qué, pudiendo oponerse a esa posesión o uso" (STC 292/2000, FJ 7).

En el contexto del Consejo de Europa, el Convenio número 108 del Consejo de Europa de 1981 -recientemente modificado por el Protocolo de Enmienda de 2018 (Convenio número 223 del Consejo de Europa)- dispone en su artículo primero que su objetivo es proteger a todas las personas físicas en lo que respecta al tratamiento de datos personales, para lo que establece una serie de garantías y limitaciones al acceso y tratamiento de los datos por terceros. En el ámbito de la Unión Europea, el artículo 16 del Tratado de Funcionamiento de la Unión Europea dispone que "Toda persona tiene derecho a la protección de los datos de carácter personal que la conciernan", y el artículo 8.1 de la Carta de los Derechos Fundamentales de la Unión Europea lo reconoce en los mismos términos[8].

Conforme el derecho fundamental a la protección de datos se ha ido viendo más comprometido a medida que se ha desarrollado el *big data*, la Unión Europea ha venido alertando de este problema y desde sus propias instituciones ha intentado adoptar soluciones para conciliar ese desarrollo con el respeto a la privacidad de las personas en este contexto. Destaca en este sentido la labor realizada por el conocido como Grupo de Trabajo del artículo 29, un órgano de naturaleza consultiva, actualmente extinto -llamado así porque fue creado al amparo de lo dispuesto por el artículo 29 de la derogada Directiva 95/46/CE relativa a la protección de datos de las personas físicas- del que pueden reseñarse diversos trabajos en los que manifestaba su preocupación por la protección de los datos personales ante los retos actuales de las nuevas tecnologías[9]. Debe destacarse, por lo que aquí interesa, su Declaración sobre el

[8] El apartado segundo dispone que "Estos datos se tratarán de modo leal, para fines concretos y sobre la base del consentimiento de la persona afectada o en virtud de otro fundamento legítimo previsto por la ley. Toda persona tiene derecho a acceder a los datos recogidos que la conciernan y a su rectificación". El apartado tercero, finalmente, establece que "El respeto de estas normas quedará sujeto al control de una autoridad independiente".

[9] Como el Dictamen 02/2013, de 27 de febrero, sobre las aplicaciones de los dispositivos inteligentes; el Dictamen 13/2011, de 16 de mayo, sobre los servicios de geolocalización en los dispositivos móviles inteligentes; la Opinión 03/2013, de 2 de abril, sobre la "limitación del propósito" (para el uso de datos, también en el *big data*); o el Dictamen 9/2011 sobre la propuesta de la industria revisada para un marco de evaluación de impacto de protección de datos y privacidad para aplicaciones RFID.

impacto del desarrollo del *big data* en la protección de las personas con respecto al procesamiento de sus datos personales en la UE[10]. Por su parte, el Parlamento Europeo había manifestado en una Resolución de 6 de julio de 2011, sobre un enfoque global de la protección de los datos personales en la Unión Europea[11], que la Directiva 95/46/CE no alcanzaba a proteger este derecho frente a los nuevos riesgos a los que se encuentra expuesto debido a los recientes desarrollos tecnológicos. En este sentido, interesan también lo dispuesto por la Comisión Europea[12], el Comité Económico y Social Europeo, y el Comité de las Regiones[13]. Se hacía necesaria una nueva regulación que garantizase la protección de este derecho y mediante la que se homogeneizasen las diversas normativas internas. Dicha norma sería el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016 (en adelante RGPD), relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos, y por el que se deroga la Directiva 95/46/CE.

4.1.2. El derecho a la protección de datos en el RGPD

El RGPD ha supuesto un importante avance para la protección de los datos personales. Entre los aspectos que pueden subrayarse, destaca el hecho de que el Reglamento haya ampliado su ámbito de aplicación, que ponga nuevos derechos a disposición de los titulares de los datos, que cree nuevas figuras ordenadas a la satisfacción de esa protección, y que introduzca una importante novedad que supone un cambio respecto de la anterior concepción del funcionamiento, por así decir, de la protección de datos, en relación con quiénes son responsables de su satisfacción. Todo ello, como no podía ser de otra manera, ha sido recogido por la Ley Orgánica 3/2018, de 5 de diciembre, de

[10] Puede consultarse en línea: https://es.calameo.com/aec/read/0018452803697a312e27b?authid=X62stTIQh47x&page=1 (última consulta: 22 de febrero de 2022).

[11] Resolución del Parlamento Europeo, de 6 de julio de 2011, sobre un enfoque global de la protección de los datos personales en la Unión Europea, en línea: https://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P7-TA-2011-0323+0+DOC+XML+V0//ES (última consulta: 22 de febrero de 2022).

[12] Comunicación de la Comisión, de 10 de junio de 2009, al Parlamento Europeo y al Consejo "Un espacio de libertad, seguridad y justicia al servicio de los ciudadanos − Programa de Estocolmo", en línea: https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52009DC0262&from=es (última consulta: 22 de febrero de 2022).

[13] Dictámenes publicados respectivamente (según dispone al principio el RGPD), en el DOC 229, de 31 de julio de 2012 (p. 90), en línea: https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=uriserv%3AOJ.C_.2012.229.01.0090.01.SPA&toc=OJ%3AC%3A2012%3A229%3AFULL, y en el DOC 391, de 18 de diciembre de 2012 (p. 127), en línea: https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=uriserv%3AOJ.C_.2012.391.01.0127.01.SPA&toc=OJ%3AC%3A2012%3A391%3ATOC (última consulta: 3 de marzo de 2022).

Protección de datos personales y garantía de los derechos digitales[14], aprobada para armonizar la legislación española con lo previsto en el Reglamento, aplicable desde mayo de ese mismo año.

En cuanto a su ámbito territorial, el Reglamento señala que "será de aplicación a quienes lleven a cabo sus actividades en la Unión, y en ese contexto traten datos personales, *aunque el tratamiento* de los datos personales *no se haga en el territorio de la Unión*" (cursiva añadida. RGPD, considerando 22).

Por lo que hace a los derechos que puede ejercer el titular de los datos personales, los clásicos derechos ARCO (acceso, rectificación, cancelación y oposición) se modifican, en parte, y se amplían con otros nuevos. El Reglamento reconoce los derechos de acceso (art. 15), rectificación (art. 16), supresión, que equivale al derecho de cancelación y que alcanza al "nuevo" derecho al olvido (art. 17), el derecho a la limitación del tratamiento (art. 18), a la portabilidad (art. 20), y a la oposición (art. 21).

Destaca la creación del Comité Europeo de Protección de Datos como organismo de la Unión (art. 68) y, a nivel interno, de los delegados de protección de datos, que deben ser designados por los responsables o encargados del tratamiento de los datos personales (art. 37). Su designación será obligatoria cuando el tratamiento de los datos personales los lleve a cabo una autoridad u organismo público (excepto los tribunales), y en un gran número de supuestos (desarrollados a nivel nacional por el art. 34 de la LOPDGDD), cuando el tratamiento de los datos lo efectúe una empresa privada. El delegado tiene atribuidas una variedad de funciones, como la de informar y asesorar al responsable o al encargado del tratamiento, y a los empleados que se ocupen del tratamiento, de las obligaciones que les incumben en relación con la protección de datos, o la de supervisar el cumplimiento de lo dispuesto en el Reglamento y las demás normas relativas a la protección de datos (art. 39).

Finalmente, como se ha señalado, se introduce una novedosa modificación en cuanto a quién se exige la satisfacción de las obligaciones impuestas por el Reglamento. La norma transfiere a los responsables del tratamiento la obligación de velar por su cumplimiento, refiriéndose en su artículo 5.2. a lo que llama "responsabilidad proactiva". Desde esta premisa, se exige la protección de datos "desde el diseño y por defecto" (art. 25), de modo que, desde el inicio -tanto en el momento de determinar los medios de tratamiento como en el momento del propio tratamiento-, la privacidad se

[14] A ésta la habían precedido la Ley Orgánica 5/1992, de 29 de octubre, reguladora del tratamiento automatizado de datos personales, conocida como LORTAD, y la Ley Orgánica 15/1999, de 5 de diciembre, de protección de datos personales, que se aprobó para trasponer al Derecho español la Directiva 95/46/CE.

integre en el tratamiento de los datos. Para ello, el RGPD pone a disposición del responsable del tratamiento (empresa privada u organismo público) medidas técnicas y organizativas para satisfacer su deber de garantizar la protección de datos personales. Se les exige, por ejemplo, que los datos estén limitados a lo necesario en relación con los fines para los que son tratados (minimización de datos) (art. 5.1.c); se prevé que, si es necesario, se proceda a la seudonimización y cifrado de los datos (art. 32) o que se realice una evaluación de impacto relativa a la protección de datos. Sobre esta evaluación, en particular, el artículo 35.1 dispone que "cuando sea probable que un tipo de tratamiento, en particular si utiliza nuevas tecnologías, por su naturaleza, alcance, contexto o fines, entrañe un alto riesgo para los derechos y libertades de las personas físicas, el responsable del tratamiento realizará, antes del tratamiento, una evaluación del impacto de las operaciones de tratamiento en la protección de datos personales".

4.1.3. Visión crítica. El limitado alcance del RGPD para la protección del derecho a la protección de datos

Sin negar los avances que el RGPD ha traído consigo, ni su voluntad de garantizar la protección de los datos personales, no pueden orillarse los muchos problemas que la norma presenta para la eficacia real de los principios que proclama y los derechos que reconoce.

Una primera cuestión clave tiene que ver con el consentimiento, en torno al que ha girado siempre el tratamiento de los datos personales. El RGPD mantiene este planteamiento, exigiendo que los datos sean tratados de forma lícita, lo que en principio sucederá si "el interesado dio su consentimiento para el tratamiento de sus datos personales para uno o varios fines específicos" (art. 6.1.a). Se ha señalado que este planteamiento está superado, pues a día de hoy no existe un control efectivo de los datos personales a través del consentimiento y los derechos a través de los que se articula. "La sociedad no está dispuesta a renunciar al uso de las IT y no tiene una fuerte cultura de la privacidad, lo que lleva a hacer casi irreal o inefectiva la garantía del consentimiento", el cual "se torna en una carta blanca al descontrol del flujo de los datos personales (...). El consentimiento acaba configurándose como un simbolismo que conlleva, a la postre, al fracaso de la privacidad pretendida y a la inoperancia del sistema de protección" (Oliver Lalana y Muñoz Soro, 2014, 163).

También puede considerarse desacertado el modo en que el Reglamento ha hecho recaer su cumplimiento en los responsables del tratamiento de los datos. A tal efecto, estos deben llevar a cabo una importante inversión en medios y preparación, así como una amplísima dedicación para el control del cumplimiento, exigencias todas que en muchos casos (pymes, autónomos) se antojan imposibles. Si a ello se añade que las nuevas previsiones normativas no

cuentan con la connivencia de las grandes corporaciones del *big data*, puede dudarse de la virtualidad del Reglamento en cuanto a que suponga una gran mejora para la protección de los datos personales desde esta perspectiva (Sancho López, 2019, 9).

Otro asunto controvertido es el hecho de que, pese a la teórica "responsabilidad proactiva" de los responsables del tratamiento, la protección de datos -en el Reglamento y también, claro, en la LOPDGDD- parece descansar asimismo, o al menos confiar, en una importante participación activa del ciudadano. Así lo acreditan la "artillería" de derechos que se le reconocen, la forma en que estos se articulan, y la realidad frente a la que están reconocidos, en la que lo habitual es que los productos que se adquieren estén preconfigurados para que la persona acepte el máximo nivel de cesión de sus datos personales, para todos los fines posibles. Pues bien, este planteamiento sería inadecuado, de un lado, porque se le presumen al ciudadano unos conocimientos técnicos de los que muy bien puede adolecer y, de otro, sobre todo, porque no parece la opción más adecuada para la protección de un derecho fundamental. Muy al contrario, de este modo se formaliza una lógica opuesta a la que subyace al sistema de garantía y protección de los derechos fundamentales propio de un Estado democrático de Derecho, y se provoca, asimismo, el indeseado *chilling effect* o efecto desaliento en los ciudadanos, quienes además ya "perciben la pérdida de privacidad como una consecuencia inevitable del progreso tecnológico, o como el precio que debemos de pagar si queremos evitar el aislamiento social" (Sancho López, 2019, 19).

Una última cuestión es relativa a la no aplicación de la normativa de protección de datos a la información anonimizada. El Reglamento dispone que "Los principios de la protección de datos deben aplicarse a toda la información relativa a una persona física identificada o identificable". Será también de aplicación a los datos personales seudonimizados, que son aquéllos que "cabría atribuir a una persona física mediante la utilización de información adicional". En cambio, tales principios "no deben aplicarse a la información anónima, es decir, información que no guarda relación con una persona física identificada o identificable, ni a los datos convertidos en anónimos de forma que el interesado no sea identificable, o deje de serlo" (considerando 26). Teóricamente la anonimización puede evitar toda posibilidad de reidentificar al individuo. Sin embargo, actualmente -se ha señalado- es imposible lograr la anonimización absoluta pues, dada la inmensa -y cada vez mayor- cantidad de datos de que dispone el *big data,* siempre existe un riesgo de que a través de los datos anonimizados se pueda reidentificar a las personas a las que están vinculados

(Morales Barceló, 2017, 6; y Gil González, 2016, 83)[15]. Qué decir, pues, de los datos que "sólo" han sido seudonimizados. Se pone en peligro de esta forma la privacidad, y el Reglamento, sencillamente, parece orillar este espinoso asunto. La LOPDGDD sí aborda algo la cuestión, aunque básicamente para garantizar que se evite la reidentificación en el ámbito del tratamiento de los datos personales de salud[16]. La norma también preceptúa como una infracción muy grave "la reversión deliberada de un procedimiento de anonimización a fin de permitir la reidentificación de los afectados" (art. 72.1 p LOPDGDD).

En relación con la obligación de realizar una evaluación de impacto en el caso de determinados tratamientos de datos (art. 35), cabría preguntarse si esta exigencia podría llegar a convertirse en una mera formalidad previa al tratamiento de los datos, cuando el informe sobre la evaluación de impacto es obligatorio.

A las críticas expuestas cabría añadir que el Reglamento abusa del uso de conceptos jurídicos indeterminados, y efectúa excesivas remisiones a las normativas internas[17]. Interesa incidir, a modo de conclusión, en el hecho de que las limitaciones que presenta el RGPD -para alzarse como una verdadera herramienta para garantizar el derecho a la protección de datos personales- tienen que ver con que la protección de datos no es el único objetivo, quizá tampoco el primero, perseguido por el Reglamento, que tiene como especial interés "asegurar el libre flujo de datos entre los Estados parte" (Sancho López, 2019, 14). El considerando segundo es clarificador en este sentido: "El presente Reglamento pretende contribuir a la plena realización de un espacio de libertad, seguridad y justicia y de una unión económica, al progreso económico y social, al refuerzo y la convergencia de las economías dentro del mercado interior, así como al bienestar de las personas físicas".

[15] Ambas autoras se encuentran citadas en la siguiente entrada del portal *Legaltoday*, en línea: https://www.legaltoday.com/practica-juridica/derecho-publico/proteccion-datos/datos-anonimizados-fuera-de-la-normativa-de-proteccion-de-datos-2020-10-01/ (última consulta: 23 de febrero de 2022). El Grupo de Trabajo del artículo 29 había advertido también sobre el riesgo de reidentificación del titular de los datos, en su Dictamen 5/2014.

[16] La disposición adicional décimo séptima de la LOPDGDD, relativa a "Tratamientos de datos de salud", prevé "una separación técnica y funcional entre el equipo investigador y quienes realicen la seudonimización y conserven la información que posibilite la reidentificación" (apartado 2, d.1), y la realización de una evaluación de impacto que "incluirá de modo específico los riesgos de reidentificación vinculados a la anonimización o seudonimización de los datos" (apartado 2, f.1).

[17] Esto se contradice con lo que cabría esperar de una norma que se adoptó para homogeneizar y superar la disparidad normativa de los Estados miembros. A nivel interno, en cuanto a la LOPDGDD, coincido con Sancho López (2019:7) en que no ha supuesto una gran mejora en relación con las dificultades que presenta el Reglamento, y en que precisamente genera más interrogantes que respuestas.

5. EL DERECHO A LA NO DISCRIMINACIÓN

5.1. Reconocimiento multinivel del derecho a la no discriminación

El artículo 14 de la Constitución española proclama la igualdad ante la ley, "sin que pueda haber discriminación alguna por razón de nacimiento, raza, sexo, religión, opinión o cualquier otra condición o circunstancia personal o social". Se trata de la igualdad formal, que determina, por un lado, que la ley ha de ser aplicada por igual a todos; que el legislador no puede dispensar injustificadamente tratos diferenciados a quienes se encuentran en situaciones jurídicas similares (igualdad en la ley); y, por otro, que los jueces han aplicar la ley por igual a quienes se encuentren en la misma situación, salvo que el cambio de criterio sea motivado y razonable (igualdad en la aplicación de la ley).

La norma, no obstante, sí puede introducir diferenciaciones entre categorías de personas que aparentemente se encuentren en la misma situación si la diferencia de trato resulta justificada de forma objetiva y razonable, y siempre que la diferenciación supere el juicio de proporcionalidad (STC 76/1990, FJ 9, STC 22/1981, FJ 3, y STC 49/1982, FJ9). Además, de esta forma (introduciendo diferenciaciones entre categorías de personas para que un colectivo o grupo pueda disfrutar como los demás del ejercicio de un derecho), el legislador satisface la igualdad material, reconocida en el artículo 9. 2 CE, según el cual "corresponde a los poderes públicos promover las condiciones para que la libertad y la igualdad del individuo y de los grupos en que se integra sean reales y efectivas".

Interesa también distinguir entre discriminación directa e indirecta, siendo la primera la que se produce cuando una persona es tratada de modo menos favorable que otra en una situación análoga a causa de su género, raza, o cualquier otra circunstancia personal o social. La discriminación indirecta, en cambio, se produce cuando una norma, actuación, o criterio aparentemente neutral genera una específica desventaja a un grupo de personas por alguna condición personal o social. El Tribunal Constitucional se pronunciaría sobre la discriminación indirecta al hilo de la discriminación por razón de sexo, afirmando que la prohibición de discriminación por razón de sexo consagrada en el art. 14 CE, "comprende no sólo la discriminación directa, es decir, el tratamiento jurídico diferenciado y desfavorable de una persona por razón de su sexo, sino también la indirecta, esto es, aquel tratamiento formalmente neutro o no discriminatorio del que se deriva, por las diversas condiciones fácticas que se dan entre trabajadores de uno y otro sexo, un impacto adverso sobre los miembros de un determinado sexo" (STC 253/2004 (FJ 7), citando la STC 198/1996, FJ 2; y, en sentido idéntico, las SSTC 145/1991, 286/1994 y 147/1995).

En el ámbito del Consejo de Europa, el artículo 14 del CEDH establece que el goce de los derechos y libertades del Convenio ha de ser asegurado sin distinción alguna, mencionándose también una serie de causas específicas de discriminación. Por lo que hace al contexto de la Unión, el artículo 21 de la Carta de los derechos fundamentales de la UE prohíbe toda discriminación, y el artículo 10 del TFUE dispone que la Unión tratará de luchar contra ella en cualquiera de sus manifestaciones. Desde la perspectiva jurisprudencial, es indudable la importancia de la labor del TEDH en la construcción de los conceptos y en la determinación del contenido de la igualdad y la no discriminación[18], aunque los pronunciamientos del TJUE en materia de discriminación indirecta parecen ser el principal referente para la jurisprudencia constitucional española[19].

Desde la perspectiva legislativa o infraconstitucional, ni a nivel nacional ni a nivel comunitario existe una única norma reguladora de la igualdad con carácter general, sino diversas normas que regulan la igualdad y la no discriminación en distintos ámbitos. En el ámbito de la Unión destacan un importante número de Directivas: sobre la no discriminación por razón del origen étnico o racial, la Directiva 2000/43/CE; sobre la no discriminación en el ámbito laboral (por razones de edad, discapacidad, religión u orientación sexual), la Directiva 2000/78/CE; y sobre no discriminación por razón de género, la Directiva 2002/73/CE (para el ámbito laboral), la Directiva 2004/113/CE, y la Directiva 2006/54/CE (también para el ámbito laboral).

En el orden interno pueden mencionarse la aprobación de la Ley 39/1999, de 5 de noviembre, para promover la conciliación de la vida familiar y laboral de las personas trabajadoras; la Ley 62/2003, de 30 de diciembre, de medidas fiscales, administrativas y del orden social, que trasponía el contenido de las dos primeras Directivas arriba citadas; la Ley Orgánica 1/2004, de 28 de diciembre, de medidas de protección integral contra la violencia de género, aprobada para erradicar esta particular forma de discriminación de la mujer; la Ley Orgánica 3/2007, del 22 de marzo, para la igualdad efectiva de mujeres y hombres, que pretendía integrar, trasponiéndolas, las Directivas 2002/73/CE y 2004/113/CE, ya mencionadas, y que se conoce como "Ley de Igualdad"; la Ley 3/2007, de 15 de marzo, reguladora de la rectificación registral de la mención

[18] Por todas, SSTEDH en los asuntos *D. H. y otros contra República Checa* (Gran Sala) núm. 57325/00, de 13 de noviembre de 2007, párrafo 175; o *Burden contra Reino Unido* (Gran Sala), núm. 13378/05, de 29 de abril de 2008, párrafo 60.

[19] Entre otras muchas, SSTJCE de 27 de junio de 1990, en los asuntos *Kowalska*; de 7 de febrero de 1991, *Nimz*; de 4 de junio de 1992, *Bötel*; o de 9 de febrero de 1999, en el *Asunto Seymour-Smith y Laura Pérez*, citadas en la STC 253/2004.

relativa al sexo de las personas; o la Ley 19/2007, de 11 de julio, contra la violencia, el racismo, la xenofobia y la intolerancia en el deporte[20].

En el contexto de la inteligencia artificial, el tratamiento de datos masivos y el uso de algoritmos puede desembocar en estos tratos discriminatorios, como ya se señaló arriba. Un derecho que paliaría la eventual discriminación derivada del tratamiento de datos y de las decisiones automatizadas sería el derecho a la transparencia algorítmica[21], que consistiría en proporcionar al interesado la información relativa al algoritmo que se ha utilizado para adoptar una decisión, como los parámetros y las concretas operaciones realizadas por el mismo. El Reglamento no reconoce la transparencia algorítmica. Sí contempla la transparencia "a secas" (artículo 12), pero orientada a garantizar la protección de datos, y no la no discriminación[22]. En efecto, es transparencia en el sentido de permitir al individuo el acceso a la información personal de la que dispone el responsable del tratamiento, y el acceso a los fines a los que está orientado su uso. Transparencia para que el ciudadano pueda procurarse y asegurarse el respeto a su privacidad.

5.2. El derecho a la no discriminación y el RGPD

Aunque, como se ha dicho, El RGPD no reconoce la transparencia algorítmica, sí contiene algunas previsiones muy importantes que podrían evitar decisiones (automatizadas) discriminatorias, contenidas en los artículos 22, 15, y en el considerando 71. El Reglamento reconoce al interesado el derecho a obtener una decisión no basada exclusivamente en el tratamiento automatizado de datos si produce en él efectos jurídicos o le afecta de modo similar (22.1), o si la decisión incluye el tratamiento de datos de las categorías sensibles del art. 9.1 RGPD (art. 22.4), salvo en algunos supuestos, como que la decisión (automatizada) sea necesaria para celebrar o ejecutar un contrato con el responsable del tratamiento de los datos, o esté autorizada por el Derecho de la Unión o de los Estados miembros, o que el interesado consienta en que la decisión sea automatizada (22.2 a, b, c). Pero, incluso en estos supuestos, el interesado tiene derecho a obtener intervención humana (22.3). Por lo demás, tiene derecho a que se le informe de la existencia de decisiones automatizadas

[20] El amplio abanico de normas que se pueden añadir a estas, de rango reglamentario y autonómico, puede consultarse en https://www.europarl.europa.eu/RegData/etudes/STUD/2020/659297/EPRS_STU(2020)659297_ES.pdf

[21] Es importante recordar en este momento que, aunque nos centremos en la discriminación, la extensión del uso de algoritmos puede provocar la vulneración de otros derechos fundamentales (derecho a la tutela judicial, el derecho de acceso a la información pública), y que la transparencia algorítmica podría servir también para solventar la lesión de esos otros derechos.

[22] Con excepción de lo dispuesto en los artículos 22.1, 13.2.f, y 15.1.h.

y, si la decisión produce en él efectos jurídicos o similares, o maneja datos de las categorías sensibles del art. 9 RGPD (o sea, las decisiones sólo "parcialmente automatizadas") tendrá, en concreto, derecho a que se le proporcione "información significativa sobre la lógica aplicada, así como la importancia y las consecuencias previstas de dicho tratamiento para el interesado" (art 15.1. h). En cambio, si se trata de decisiones "automatizadas puras" (art. 22.3), la interpretación del artículo 22.3 en conexión con lo dispuesto en el considerando 71 *in fine* determina que en estos supuestos habrá de proporcionarse la información específica al interesado, que tendrá derecho a obtener intervención humana, a expresar su punto de vista, y a recibir una *explicación* de la decisión tomada.

Otro precepto del RGPD que interesa desde la perspectiva de la no discriminación es su artículo 9.1, que prohíbe "el tratamiento de datos personales que revelen el origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas, o la afiliación sindical, y el tratamiento de datos genéticos, datos biométricos dirigidos a identificar de manera unívoca a una persona física, datos relativos a la salud o datos relativos a la vida sexual o las orientación sexuales de una persona física". El precepto proscribe el tratamiento de estos datos salvo que concurra alguna de las circunstancias del apartado segundo, como, por ejemplo, que el interesado haya dado su consentimiento explícito (a), que el tratamiento sea necesario para proteger intereses vitales del interesado o de otra persona física (c), o que el tratamiento se refiera a datos personales que el interesado ha hecho manifiestamente públicos (e). La especial protección de estos datos se debe a que el uso de alguna de estas categorías es automáticamente sospechoso de ser discriminatorio, pues históricamente -como ya se ha visto- han sido motivo de discriminación. La novedad que presenta el RGPD en relación con la protección de las categorías especiales de datos personales es la inclusión en ellas de los datos genéticos y los biométricos (que no aparecían en la Directiva 95/46/CE)[23].

La discriminación puede producirse, como ya se indicó, en cualquier momento del proceso, desde que los datos son recogidos hasta la aplicación del algoritmo por el que se toma la decisión. Interesa en este sentido la previsión del RGPD relativa a la obligación de realizar evaluaciones de impacto previamente a determinados tratamientos de datos, ya mencionada, del artículo 35. Con esta evaluación del impacto, los responsables del tratamiento podrían detectar, y prevenir, un riesgo de discriminación en el tratamiento de los datos y/o en el algoritmo a través del que se adopte una decisión automatizada.

[23] La LOPDGDD se refiere también a las categorías especiales de datos (a algunas de ellas) en su artículo 9.

Concretamente, el apartado segundo del artículo 35 establece que la evaluación de impacto relativa a la protección de los datos a que se refiere el apartado primero se requerirá, entre otros, "en caso de evaluación sistemática y exhaustiva de aspectos personales de personas físicas que se base en un tratamiento automatizado, como la elaboración de perfiles, y sobre cuya base se tomen decisiones que produzcan efectos jurídicos para las personas físicas o que les afecten significativamente de modo similar" (a), y de "tratamiento a gran escala de las categorías especiales de datos a que se refiere el artículo" (b).

5.3. Apuntes críticos. Limitaciones del RGPD en relación con el derecho a la no discriminación

Lo primero que ha de señalarse es la ausencia de reconocimiento del derecho a la transparencia algorítmica en el Reglamento. Desde tiempo antes de la adopción de esta norma se había venido alertando del riesgo de que la extensión del uso de decisiones automáticas y análisis predictivos pudiese conducir a cambios indeseables en el desarrollo de nuestras sociedades, dirigiéndolas hacia la discriminación, el refuerzo de estereotipos ya existentes, la segregación social y cultural, y la exclusión[24]. Sin su inclusión en el Reglamento parece haberse perdido una gran oportunidad para el reconocimiento de este derecho, el mejor instrumento para que el interesado pueda cerciorarse de que no se le ha discriminado cuando se ha tomado una decisión automatizada (o semiautomatizada), pero también para la protección de otros intereses reglamentaria y constitucionalmente relevantes.

El reconocimiento en el Reglamento del derecho a que no se adopten decisiones individuales automatizadas es, en mi opinión, confuso. Se trata de un derecho complejo, con demasiadas excepciones, y que deja abiertas cuestiones importantes. ¿Hasta dónde alcanza el "derecho a la explicación" que puede deducirse del artículo 22.3 en conexión con el considerando 71? ¿Puede considerarse equivalente al derecho a la transparencia algorítmica? Por otro lado, ¿qué sucede con las decisiones en las que haya habido una intervención humana "menor", en el sentido de que básicamente ha sido una decisión algorítmica que una persona se ha limitado a plasmar? Parece que en estos casos no existiría el derecho a la explicación, en lo que quiera que consista.

Hasta aquí se ha abordado la cuestión de la discriminación desde una perspectiva individualizada, con referencia a garantías como la información de la lógica aplicada en la decisión que afecte al interesado, o el derecho a una

[24] En este sentido se pronuncia el Supervisor Europeo para la Protección de Datos, en su Opinión 7/2015, *Meeting the challenges of big data A call for transparency, user control, data protection by design and accountability*, en línea: https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf (última consulta: 23 de febrero de 2022).

explicación sobre la misma. Algo puede decirse también en este apartado sobre las previsiones normativas que podrían proteger frente a la discriminación en los momentos anteriores del proceso de tratamiento de los datos. Concretamente, sobre la obligación de realizar una evaluación de impacto en el caso de determinados tratamientos de datos (art. 35 RGPD), más allá de las dudas sobre su virtualidad ya manifestadas al hilo del derecho a la protección de datos, cabría preguntarse si esta herramienta realmente puede servir para apreciar la posible discriminación, pues, como se señaló, en ocasiones el sesgo será difícil de detectar y la evaluación de impacto no está específicamente orientada a descubrir este tipo de discriminación.

6. APUNTES PROPOSITIVOS (Y CONCLUSIVOS)

A la vista de los problemas que presenta la legislación para garantizar una efectiva protección de los derechos incididos, pueden realizarse, a modo de conclusión, algunas reflexiones, principalmente sobre las propuestas que se han planteado para paliar estas dificultades.

1. Una primera cuestión tiene que ver con la importancia de la ética en este ámbito. Habida cuenta de la incapacidad de ciertos planteamientos clásicos para la protección de los derechos incididos, como hemos señalado en relación con el consentimiento y el derecho a la protección de datos, se ha reivindicado insistentemente la importancia de los principios éticos en este contexto, no sólo por parte de la doctrina científica, sino desde las instituciones más relevantes. Tal es el caso del Parlamento Europeo, que señala, en su Resolución de 2017 sobre las implicaciones de los macrodatos en los derechos fundamentales, que "son fundamentales normas científicas y éticas estrictas para gestionar la recopilación de datos y valorar los resultados" de los análisis basados en macrodatos (consideración segunda); y se refiere a la necesidad de que los Estados miembros "desarrollen un marco ético común sólido para el tratamiento transparente de los datos personales y la toma de decisiones automatizada que sirva de guía para la utilización de los datos y la aplicación en curso del Derecho de la Unión" (consideración vigésima). Claro que, a pesar de la importancia de establecer unos principios éticos fuertes en el ámbito de la inteligencia artificial, la ética no puede sustituir al Derecho. Derecho, además, en el que entiendo que ha de primar la heterorregulación frente a la autorregulación. En este sentido, la LOPDGDD prevé que los códigos de conducta, cuya adhesión es potestativa, sean complementarios.

2. Dada la incapacidad de ciertos derechos, o de sus versiones clásicas, para la protección de los intereses de las personas frente a las nuevas tecnologías, se ha planteado la importancia de reflexionar sobre la conveniencia de reconocer nuevos derechos en el contexto de la inteligencia artificial, que pudiesen incardinarse en su caso, entiendo, en el contenido del pertinente

derecho fundamental. Se habla, por ejemplo, de un "derecho a la criptografía" (Cotino Hueso, 2017, 137), como una nueva versión del derecho a la privacidad. Ejemplos de derechos nuevos en este contexto son el derecho al olvido (de naturaleza jurisprudencial hasta su reconocimiento por el RGPD) o el derecho a no ser objeto de una decisión "automatizada pura" en determinados supuestos.

3. Ante las limitaciones de las garantías subjetivas, por las dificultades de ejercer exitosamente un control de los datos personales, deben ganar importancia las garantías objetivas (Cotino Hueso, 2017, 146), de carácter preventivo, que se articulen en forma de obligaciones sobre los responsables del tratamiento de los datos. El RGPD establece, en este sentido, el respeto de la privacidad de los datos desde el diseño y por defecto, así como la obligación de efectuar una evaluación de impacto, o la de minimizar el uso de datos personales, entre otras. Para garantizar el cumplimiento de estas obligaciones, los poderes públicos deben buscar fórmulas para evitar que la adopción de las medidas sea excesivamente gravosa para los responsables del tratamiento y, a la vez, deben de ser muy estrictos en sus funciones de inspección y auditorías (que se atribuyen a las autoridades de control, que en España es la AEPD). Auditorías, por cierto, que deberían ser unas para la protección de datos y otras específicas para la protección frente a la no discriminación, especialmente enfocadas, entre otros aspectos, a la detección de sesgos algorítmicos.

4. En particular, en relación con el derecho a la protección de datos, un elemento que cabría introducir para mejorar su protección sería el establecimiento de una limitación temporal real -o, al menos, más efectiva- para la conservación de los datos. El artículo 14.2 RGPD dispone que el responsable del tratamiento facilitará al interesado la siguiente información necesaria para garantizar un tratamiento de datos leal y transparente: "a) el plazo durante el cual se conservarán los datos personales o, cuando eso no sea posible, los criterios utilizados para determinar este plazo". Por supuesto, las empresas recurren siempre a esta segunda opción ("los criterios"). La LOPDGDD tampoco ha solucionado el problema. La cuestión de los plazos es especialmente sangrante en relación con el derecho al olvido en las búsquedas de internet. Si existe este derecho, cabe preguntarse si no debería obligarse a los motores de búsqueda a suprimir la información de las personas cuando devengan las circunstancias del artículo 93 RGPD o cuando pase cierto tiempo.

5. Debe ponerse en valor, a modo de conclusión, la importante labor jurisprudencial que está siendo llevada a cabo por algunos tribunales. En relación con el derecho a la no discriminación, existen novedosas decisiones judiciales que han hecho valer este derecho cuando ha resultado lesionado por el uso de un algoritmo discriminatorio, tanto frente al poder público como frente a los particulares (en el ámbito laboral). En relación con lo primero, tuvo

bastante repercusión una sentencia holandesa, dictada en el asunto *Siry* (*System Risk Indication*) por el Tribunal de Distrito de La Haya el 5 de febrero de 2020. En ella se declararía ilegal un sistema algorítmico utilizado por el Gobierno holandés con el que se trataba de prevenir y combatir el fraude fiscal y a la seguridad social. Al aplicarse sólo en determinadas zonas integradas por barrios cuya población se consideraba que tenía una mayor predisposición a este tipo de fraude, el sistema algorítmico lesionaba el derecho a la igualdad de trato de la población más pobre e inmigrante. Esta sentencia, además, estimaba que se producía también lesión del derecho a la privacidad del art. 8 CEDH, pues se habían utilizado una cantidad ingente de datos personales, de forma desproporcionada y con incumplimiento de los principios de limitación de la finalidad y minimización (uso de los datos necesarios) (*The technolawgist*, 2020).

Por lo que hace a la eficacia horizontal del derecho a la no discriminación, destaca una sentencia italiana, dictada por el Tribunal ordinario de Bolonia, en la que la magistrada concluiría que el algoritmo utilizado por la empresa Deliveroo para organizar los horarios de los repartidores (*riders*) era discriminatorio. El repartidor podía reservar un tramo horario de reparto (en función de unos órdenes preestablecidos de selección de los mismos). Si anulaba el tramo elegido se le penalizaba y perdía su lugar en el turno de elección. La penalización era aún mayor si no lo anulaba y no acudía a repartir. Esto era así independientemente del motivo por el que lo hiciese, que sencillamente no podía alegarse, lo que, a juicio de la magistrada, suponía una discriminación indirecta: "El sistema de perfilación del *rider* adoptado por la plataforma Deliveroo, basado sobre los dos parámetros de la reputación o confianza y la participación, al tratar del mismo modo a quien no participa por motivos fútiles y a quien no participa porque hace huelga (o porque está enfermo, o lesionado, o asiste a un menor enfermo, etc), discrimina en concreto a este último, marginándolo del grupo prioritario y por lo tanto reduciendo significativamente sus futuras ocasiones de acceso al trabajo" (Baylos Grau, 2021).

6. Algunos tribunales, por tanto, están marcando pautas interesantes para el respeto de los derechos aquí incididos. Por lo que hace al derecho a la protección de datos, la labor judicial se ha desarrollado al máximo nivel, pues es el Tribunal Superior de Justicia de la Unión el que nos ha dejado ya relevantes sentencias para garantizar su protección. No puede dejar de mencionarse el *caso Costeja* (2014). Desde esta polémica decisión (STJUE en el asunto *Google contra AEPD*), bien conocida, se reconoce el derecho al olvido frente a los motores de búsqueda en internet. Otras dos importantes decisiones del TJUE son las adoptadas en el caso *Schrems (I y II)* (2015 y 2020) en que se cuestiona, a raíz de las revelaciones del analista Snowden sobre la inteligencia norteamericana, la existencia de un adecuado nivel de protección de los datos personales transferidos a ese país. En ambas sentencias -en las que invalida los

sistemas que daban cobertura a la transferencia de datos personales desde la UE a EEUU, conocidos como *Safe Harbour* (2015) y *Privacy Shield* (2020)- el TJUE sostiene que la comunicación de los datos personales debe contar con el consentimiento del afectado o tener un fundamento legalmente previsto (art. 8 CDFUE), y que cualquier normativa que autorice el acceso *generalizado* a los datos personales sin establecer ninguna salvaguarda vulnera el contenido esencial del derecho a la vida privada y familiar (art. 7 CDFUE) (Fuentes Máiquez, 2020).

Finalmente, también en el ámbito de la jurisprudencia constitucional española debe destacarse algún pronunciamiento. En concreto la STC 76/2019, que declaró inconstitucional la habilitación legal que permitía a los partidos la elaboración de perfiles políticos de los ciudadanos para personalizar los mensajes de propaganda. Fue el Defensor del pueblo quien interpuso el recurso de inconstitucionalidad contra el nuevo artículo 58 bis de la Ley Orgánica del Régimen Electoral General (LOREG), introducido -precisamente- por la disposición final tercera, apartado dos, de la Ley Orgánica 3/2018, de 5 de diciembre, de protección de datos personales y garantía de los derechos digitales. El Tribunal Constitucional consideró, en efecto, que ese artículo que posibilitaba el *profiling* político era lesivo del derecho a la protección de datos del artículo 18.4 CE.

7. BIBLIOGRAFÍA

AEPD - ISMS Forum (eds.), Aced, Emilio *et al.* (coords.) (2017), *Código de buenas prácticas en protección de datos para proyectos de Big Data*, AEPD e ISMS Forum, Madrid.

Arellano Toledo, Wilma, (2019), "El derecho a la transparencia algorítimica en *big data* e inteligencia artificial", en: *Revista General de Derecho Administrativo*, 50, 1-28.

Baylos Grau, Antonio (2021), "El algoritmo no es neutral. No permite ejercer derechos fundamentales a los trabajadores de plataformas", en: *Blog de Antonio Baylos*, disponible en línea: https://baylos.blogspot.com/2021/01/el-algoritmo-no-es-neutral-no-permite.html (última consulta: 2 de marzo de 2022).

Castellanos Claramunt, Jorge, y Montero Caro, María Dolores (2020), "Perspectiva constitucional de las garantías de aplicación de la inteligencia artificial: la ineludible protección de los derechos fundamentales", en: *Ius et Scientia*, 6, 2, 72-82.

Cotino Hueso, Lorenzo (2017), "*Big data* e inteligencia artificial, una aproximación a su tratamiento jurídico desde los derechos fundamentales", en: *Dilemata*, 24, 131-150.

— (2019), "Riesgos e impactos del *big data*, la inteligencia artificial y la robótica. Enfoques, modelos y principios de la respuesta del Derecho", en: *Revista General de Derecho Administrativo*, 50, 1-37.

Eguíluz Castañeira, Josu Andoni (2020), "Desafíos y retos que plantean las decisiones automatizadas y los perfilados para los derechos fundamentales", en: *Estudios de Deusto*, 68, 2, 325-368.

Fuentes Maíquez, Antonio (2020), "Comentario de la STJUE de 16 de Julio de 2020, C-311/18 (Schrems II)", en: *Icade: Revista de la Facultad de Derecho*, 110, 1-10.

García Mahamud, Rosario (2015), "Partidos políticos y derecho a la protección de datos en campaña electoral: tensiones y conflictos en el ordenamiento español", en: *Teoría y Realidad Constitucional*, 35, 309-338.

Gil González, Elena, 2016, *Big data, privacidad y protección de datos*, Agencia Estatal de Protección de Datos-Boletín Oficial del Estado, Madrid.

Han, Byung-Chul (2014), *Psicopolítica, Neoliberalismo y nuevas técnicas de poder*, Herder, Barcelona.

Martínez, Ricard (2014): "Ética y privacidad de los datos", en: *Big Data: de la investigación científica a la gestión empresarial* (jornadas), Fundación Ramón Areces, 3 de julio de 2014, disponible en línea: http://sgfm.elcorteingles.es/SGFM/FRA/recursos/conferencias/ppt/1776180509_1472014102438.docx (22/2/2022).

Medina Guerrero, Manuel (2021), "De algoritmos y otras palabras inquietantes", en: *Blog de Asociación de Constitucionalistas de España*, disponible en línea: https://www.acoes.es/de-algoritmos-y-otras-palabras-inquietantes/ (1/3/2022).

Morales Barceló, Judith (2017), "Big Data y Protección de Datos: especial referencia al consentimiento del afectado", en: *Revista Aranzadi de Derecho y Nuevas Tecnologías*, 44, 137-164.

Moreno Muñoz, Miguel (2017), "Privacidad y procesado automático de datos personales mediante aplicaciones y bots", en: *Dilemata*, 24, 1-23.

Oliver Lalana, Daniel y Muñoz Soro, José Félix (2014): "El mito del consentimiento, o por qué un sistema individualista de protección de datos (ya) no sirve para (casi) nada", en: Valero Torrijos, Julián (coord.), *La protección de los datos personales en Internet ante la innovación tecnológica*, Aranzadi, Cizur Menor, 153-196.

Sancho López, Marina (2019), "Estrategias legales para garantizar los derechos fundamentales frente a los desafíos del *big data*", en: *Revista General de Derecho Administrativo*, 50, 1-28.

Thetechnolawgist (2020), "Derechos humanos en un mundo de algoritmos. La sentencia histórica que ata en corto la implantación de modelos opacos", en el portal *Thetechnolawgist.com*, disponible en línea: https://www.thetechnolawgist.com/2020/02/12/derechos-humanos-en-un-mundo-de-algoritmos-la-sentencia-historica-que-ata-en-corto-la-implantacion-de-modelos-opacos/ (última consulta: 3 de marzo de 2022).

Zuiderveen Borgesius, Frederik (2018), *Discrimination, artificial intelligence, and algorithmic decision-making*, Consejo de Europa, Estrasburgo.

CAPÍTULO V

LA FRAGILIDAD DE LA VERDAD EN LA SOCIEDAD DIGITAL[1]

Mª OLGA SÁNCHEZ MARTÍNEZ
Universidad de Cantabria
maria.sanchez@unican.es

1. INTRODUCCIÓN

Las comunicaciones tienen, sin lugar a dudas, un gran protagonismo en las sociedades del siglo XXI. Internet y las redes sociales son los medios de comunicación por excelencia, el hábitat de la información y también de la desinformación. Siendo la información un elemento imprescindible para el buen funcionamiento del sistema democrático, la desinformación afectará negativamente a la democracia y al eficaz ejercicio de algunos de sus derechos.

Si bien la desinformación ha existido siempre, en la era digital adquiere perfiles propios, acogiendo una nueva causa de la misma: la posverdad. De la posverdad se dice que genera algunos rasgos culturales específicos que, teniendo como centro la pérdida de valor de la verdad y la prioridad de lo emocional sobre lo racional, puede llegar a comprometer el pluralismo, el debate y el acuerdo. Valores propios del sistema democrático que se resienten por algunos efectos perversos de un uso inadecuado de las comunicaciones digitales, como el escepticismo, la simplificación, la radicalización y la polarización de las ideas.

2. INTERNET Y SUS POSIBILIDADES PARA MEJORAR LA DEMOCRACIA

Las nuevas tecnologías de la información y comunicación, especialmente la llegada y el uso generalizado de internet, han supuesto un profundo cambio en los modos de vida de las sociedades del siglo XXI. Cambios que afectan a todos los ámbitos, desde el ocio y la cultura, pasando por las relaciones sociales, la economía y la política. En lo referente a la política, la banda ancha fue saludada con gran entusiasmo y optimismo, como un mecanismo para la promoción de la democracia. Se confiaba que internet podía contribuir de manera decisiva a eliminar las barreras informativas y, con ello, hacer insostenibles los sistemas autoritarios (Levy, 2002, 50; Fukuyama, 2003, 21).

[1] Este trabajo se ha realizado en el marco del proyecto de investigación "La inteligencia artificial jurídica" (RTI2018-096601-B-100 MCIU/AEI/FEDER, UE), del Programa Estatal de I+D+i Orientada a los Retos de la Sociedad.

La red incrementa sustancialmente las posibilidades de comunicación de cualquier ciudadano y ciudadana, desde cualquier lugar del mundo. Un pequeño dispositivo electrónico, al alcance prácticamente de cualquier persona, permite una rápida y extensa difusión de los mensajes. Circunstancias que sitúan a la información en unas nuevas coordenadas, con indudables efectos en el desarrollo de la vida social y política, y en la formación de la opinión pública. Este incremento del potencial comunicativo en la sociedad digital es un activo para procurar condiciones adecuadas que mejoren la democratización de la vida en general, y de la vida política en particular.

El sistema democrático es un procedimiento de toma de decisiones que, partiendo de la confrontación de ideas y opiniones, del diálogo y el debate, permite llegar a acuerdos para decidir sobre distintas opciones, dirigidas a ordenar la vida en común de la ciudadanía. Ahora bien, la democracia, más allá de un procedimiento para la toma de decisiones, es un reconocimiento del valor de las personas y sus derechos, de quienes de forma autónoma intervienen decisivamente en dicho procedimiento. No hay democracia sin reconocimiento de los valores de libertad, igualdad, pluralismo, autonomía y de los principios y derechos que los hacen posible. Pues bien, la tecnología digital ofrece posibilidades de actuación que permiten mejorar la efectividad de aquellos valores, principios y derechos, siendo capaz de proporcionar a ciudadanos y ciudadanas un protagonismo en el funcionamiento del sistema democrático inédito hasta el momento.

Dichas posibilidades tienen que ver con el acceso a la información y la libre expresión del pensamiento y manifestación de la opinión, elementos básicos y estratégicos para un correcto funcionamiento de la democracia. Hasta la era internet, la recepción y transmisión de información, así como la libre expresión del pensamientos e ideas, se enmarcaban principalmente en un proceso de intermediación, en manos de profesionales y empresarios de los medido de comunicación. Actualmente, es posible que la expresión, opinión e información fluyan sin necesidad de intermediarios. En la red, ciudadanos y ciudadanas no somos meros consumidores de información, sino usuarios "con poderes propios" (Rheingold, 2004, 223). Cualquier persona, sin muchos recursos, puede generar su propia información, dar su opinión y llegar a un gran número de personas sobre las que ejercer influencia y, por tanto, tener la oportunidad de contribuir activamente a conformar opinión pública. En este sentido, y como consecuencia, valores y derechos, como la libertad y la igualdad pueden resultar reforzados y, por ello, internet puede tener un gran potencial en el desarrollo de las democracias.

Ahora bien, estas nuevas coordenadas de libertad e igualdad se desenvuelven en un contexto que tiene sus propias peculiaridades, no

necesariamente favorables, sin fisuras, a la libertad e igualdad, ni siempre soportes sólidos para el pluralismo, ni tampoco se sostienen, sin condiciones, en un clima estable en el que generar los consensos necesarios en una sociedad democrática. La llegada de la sociedad digital ha supuesto la ruptura de algunos vínculos jerárquicos tradicionales, pero también la creación de nuevas jerarquías de control y poder. Como consecuencia, se han producido algunos cambios en los métodos para asignar prioridades en el ámbito público, que pueden ser determinantes para la toma de decisiones políticas y para la redistribución de roles de poder (Mathias, 1998, 26, 30, 44).

En el momento actual, el optimismo que invadió la llegada de internet, en relación a la promoción de la democracia y la derrota del autoritarismo, no puede considerarse un objetivo realizado. Internet puede ser una herramienta de libertad, pero también de opresión. La libertad de disfrutar de internet, si existe, no significa necesariamente un incremento de la libertad de la ciudadanía. Internet es también un instrumento de vigilancia, de propaganda y de censura (Morozov, 2012, 124). Inmersos en un panóptico digital asistimos a una exposición pública de nuestras vidas sin precedentes, a una radical transparencia de muchos aspectos -tanto del ámbito privado, como público-, a la pérdida de secretos y misterios. Con ello, se consigue proyectar una mirada sentida, en muchas ocasiones, como despótica y, entonces, resultar un obstáculo, más que un aliciente, para la comunicación. De forma muy ilustrativa se ha denominado a este efecto, poco proclive a impulsar una mayor libertad, como la "dictadura de la hipervisibilidad" (Cardon, 2012, 212).

Pocas dudas se tienen en este momento de que las tecnologías digitales, como forma de comunicación e información, aportan muchos beneficios y algunos riesgos. Paradójicamente, y pese a su potencial para lo contrario, pueden contribuir a erosionar el pensamiento crítico, por su tendencia a eliminar de nuestro entorno la disidencia, el desacuerdo o la confrontación. En lugar de ahondar en el pluralismo, podrían hacerlo en la uniformidad, crear una falsa sensación de consenso y eludir la complejidad, facilitando que nos acomodemos en lo superficial. La amplitud del espacio virtual no genera, necesariamente, un ámbito común de comunicación. Por el contrario, puede propiciar la creación de caminos paralelos, por los que circular agrupados según se compartan las mismas ideas, y sin un lugar de encuentro con el diferente. En estos casos, internet no es adecuado para mejorar los métodos y procedimientos propios de la democracia.

Al respecto, el papel de la verdad, tan importante para recibir información, conformar la opinión y manifestarla al amparo de la libertad de expresión, se encuentra sometida en la sociedad digital a nuevas tensiones, que

afectarán a la calidad de las instituciones democráticas y los derechos de la ciudadanía.

3. EL DOBLE EFECTO DE LA RED SOBRE LA INFORMACIÓN: EXCESOS Y DEFECTOS

Las primeras evidencias de los efectos del desarrollo de la tecnología digital e Internet se refieren a la ingente cantidad de información que prolifera en la red y la celeridad de su circulación. De un lado, el contexto digital ofrece enormes posibilidades y facilidades para crear y difundir información que pueda incrementar los saberes de unos ciudadanos y ciudadanas que, informados y formados en esta nueva era digital, Mason denomina "cultos universales" (Mason, 2016, 162-163). Es innegable que contar con una información amplia y veraz es fundamental para formar las ideas y preferencias de la ciudadanía, para argumentar y contraargumentar, en definitiva, para configurar los elementos propios del debate inherente a un sistema democrático.

Ahora bien, si como aspecto positivo, la información en el contexto digital puede mejorar el sistema democrático, la desinformación habrá de afectar negativamente a la democracia. Por ello, a lo largo de la historia cada avance tecnológico facilitador de la circulación informativa, ha venido acompañado de un incremento de los riesgos asociados a la misma. En el momento actual, las inmensas potencialidades de internet para facilitar la producción y difusión de noticias, su gran rentabilidad económica, la carga emocional que comporta y los efectos políticos que puede suponer, han multiplicado la sensación de los riesgos capaces de producir los contenidos en red. Las tecnologías de la comunicación digital tienen un enorme potencial para informar, pero también para desinformar, aún más, para manipular, falsear, saturar y colapsar la información, con todo lo perverso que esto acarrea para la democracia (Hernández Pérez, 2018, 204-206; Alonso González, 2019, 32).

El primer riesgo se percibe atendiendo a un criterio puramente cuantitativo. El exceso de noticias es inherente al contexto tecnológico digital y sus consecuencias no son difíciles de constatar: en el mejor de los casos un "despilfarro informativo"; en el peor, una gran dificultad para discriminar lo importante, lo conveniente o lo interesante (Rheingold, 2004, 223). Se ha señalado al respecto que la cantidad y velocidad de información circulando por la red genera una "fatiga infinita y desproporcionada", en lugar de posibilidades de elección múltiples (Sartori, 1998, 135). Internet ha sido definida como "una máquina diseñada para la recogida, transmisión y manipulación eficiente y automatizada de información" (Carr, 2011, 184; Pariser, 2017, 137). Una suerte de contaminación informativa que exigiría "inventar una dietética de la información" (Rosnay, 1996, 243). Pero será necesario advertir sobre las

"dietas informativas" no saludables, porque del "empacho" de información podemos pasar a una "malnutrición" o a un "determinismo informativo". Una "malnutrición" informativa con la que se corre el riesgo de ser sometidos a una "fatiga" que llegue a provocar rechazo y casi "asfixia" (Sartori, 1998, 130, 135).

El extraordinario incremento de contenidos informativos en internet afecta directamente a la posibilidad y capacidad de procesarlos, de entenderlos, contrastarlos, comprobar su veracidad y establecer diferencias sobre su interés, trascendencia o importancia. Se ha señalado que la información excesiva puede producir confusión y frustración, también una visión borrosa y falsa (Gleick, 2011, 428). Así pues, la misma tecnología que lleva aparejada la abundancia de información, paradójicamente, puede producir también una escasez informativa. A veces, la falta de información no se debe a las propias limitaciones del ser humano para enfrentarse a los inmensos contenidos disponibles en internet, sino porque se condiciona o limita la información que se recibe, a base de filtros aplicados por programas informáticos, que toman como referencia los rastros que dejamos cada vez que nos adentramos en la red.

En cualquier caso, ni los excesos, ni los defectos de información, resultan inocuos para los ciudadanos y ciudadanas receptoras de la misma, para las sociedades que habitamos y para el desarrollo de los sistemas democráticos que las gestionan. Uno de los primeros efectos de las disfuncionalidades en la información es la pérdida de credibilidad y de confianza en la misma, acompañado por el escepticismo instalado en la ciudadanía (García Fernández, 2007, 80-87). Pero no sólo eso, la rapidez de la circulación de las noticias repercute en la reducción del tiempo del que se dispone para contrastarlas, someterlas a reflexión y crítica, limitando las reacciones a un sencillo "me gusta" o un "no me gusta". Una reducción que dificulta conectar con pensamientos y emociones más complejas y comprometidas con aquello que nos rodea y que termina por aceptar una equiparación entre lo más valioso y lo más visto (Zafra, 2017, 184-186). Se aboca así al fomento de la versión más simplificada del ciudadano. La visión del internauta como un "homo communicans", puede perderse para quedar reducido a un simple contacto (Han, 2014, 89) y el "homo sapiens" involucionar a "homo insipiens", transmisor y receptor de recipientes vacíos (Sartori, 1998, 96-97, 145-146). La ciudadanía en esta versión simplificada puede adoptar también una perspectiva simple de la realidad, tendente a centrar la atención en generalidades y arrinconar lo particular, aquello que permite establecer matices. Siendo poco permeables a recibir ideas nuevas, se limita la diversidad, para concluir con la polarización de las ideas propias y su radicalización en el contexto social.

Escepticismo, reducción, simplicidad, polarización y radicalidad son algunos de los riesgos de las sociedades democráticas, que se han acentuado en

la sociedad tecnológica. Factores que no invitan a un optimismo sin paliativos para nuestro futuro personal, social y político más inmediato. Pero tampoco han de conducir necesariamente al pesimismo. La realidad es que, cómo muy bien se ha expresado, internet es un "expositor y un vertedero" que contiene "la joya y la basura" (Serna, 2017, 115-116). Explorar sus efectos positivos -sus joyas- y tratar de neutralizar o mitigar los negativos -sus residuos no reciclables- es una tarea necesaria para aprovechar el gran potencial de las nuevas tecnologías y ahondar en la mejora de nuestras sociedades y de quienes en ellas vivimos.

4. LA VERDAD COMPROMETIDA POR LA TECNOLOGÍA

La facilidad de la tecnología digital para crear y difundir información alcanza tanto a aquella que es verdadera, como a la que es falsa. Sus procedimientos se encuentran bastante alejados de los parámetros de la situación ideal de habla que propone Habermas, para que la verdad pueda significar la promesa de alcanzar un consenso racional sobre lo dicho (Habermas, 1998, 121, 150-158). En la red cabe cualquier contenido: información, desinformación, opinión, creencia, verdad, falsedad. En consecuencia, puede tratarse como una herramienta útil para formar e informar. Pero, de la misma manera, pude ser aprovechada para deformar y desinformar.

Ahora bien, sería injusto señalar a la tecnología digital y a las redes sociales como únicas responsables de las noticias falsas y la desinformación que comportan. Las campañas de desinformación, las noticias falsas, la propaganda y las mentiras han existido siempre, y los avances tecnológicos ha supuesto, en cualquier momento de la historia, un incremento del riesgo a su expansión. Los temores planteados con la aparición de la imprenta son un buen ejemplo de ello. Frente a la tradición oral y los manuscritos, las obras impresas supusieron una revolución para la difusión de textos. La imprenta, considerada como uno de los inventos que cambiaron la apariencia y el estado del mundo entero (Bacon, 2011) permitió el acceso, en periodos cortos de tiempo, de multitud de personas a todo tipo de producción escrita (Einstein, 1994, Id., 2010). Sus detractores pusieron de manifiesto los riesgos que entrañarían la divulgación masiva de los textos, en orden a dificultar su comprensión y capacidad de crítica; así como la imposibilidad de controlar su calidad, detectar los posibles escritos engañosos y la facilidad para propagar noticias falsas (Eisenstein, 1994; Id., 2010)[2].

[2] No faltaron tampoco reticencias referidas a la posibilidad del incremento de daños en el honor de las personas, como consecuencia de la producción masiva de textos. Incluso se temieron las consecuencias que para el poder tenía la libertad que pudiera conferir a la ciudadanía el saber transmitido por este medio de difusión. Al principio del segundo acto de la obra de Lope de Vega, *Fuenteovejuna*, la conversación entre el licenciado Barrildo y Leonelo refleja un debate en

(...)

Si bien es cierto que, no siendo una novedad, se ha ido modificando la percepción y la intensidad de la dimensión de los peligros que constituyen las tecnologías de la comunicación, especialmente cuando se compromete la veracidad o cuando sus contenidos afectan a grupos vulnerables. No es difícil detectar que las redes sociales y las plataformas virtuales, con sus mensajes cortos, la facilidad de difusión, su apelación a lo emocional, su recurrente finalidad más persuasiva que informativa, son un ecosistema de las noticias verdaderas y también de aquellas que no lo son, de las *fake news*, de la verdad a la carta, de la posverdad y de la mentira. Hasta tal punto, que la desinformación en la era digital, sostenida en la falsedad, la propaganda y la pseudociencia, se ha convertido en una materia de negocio rentable (D´Ancona, 2019, 57) y las noticias falsas, a tenor de alguna opinión, en una "epidemia", una vulneración del derecho a la información y un límite a la libertad de expresión, consideradas en el Foro contra las *fake news* organizado por el Parlamento Europeo, celebrado en Madrid el 8 de mayo de 2018, como una auténtica "amenaza" contra la democracia y sus libertades (Richter Morales, 2018, 30, 42).

Ya se ha señalado que la información es un elemento funcional imprescindible para la formación de la opinión pública y necesaria para el correcto desenvolvimiento de la democracia. Siendo así, es lógico que la desinformación produzca un efecto desestabilizador del sistema. A veces, la desinformación se deriva de errores o equivocaciones, propios de la rapidez en que la información se desarrolla en el entorno digital, y, en tal contexto, la quiebra de la capacidad de contraste, reflexión y precisión. Pero, en otras ocasiones, la desinformación proviene, y puede que no en menor medida, de comportamientos intencionados y premeditados, con propósito de dañar y deformar a la opinión pública. Los efectos, en este caso, resultarán aún más perniciosos para el sistema democrático. Se trata de noticia falsas deliberadamente creadas con el objetivo, por un lado, de obtener beneficios económicos, porque la manipulación informativa es un negocio ventajoso, a través del que se incentivan las visitas en determinadas páginas web, y así poder obtener notables ingresos en publicidad. Por otro lado, conseguir ventaja política, tratando de influir en la opinión pública, para apoyar o desacreditar determinadas posiciones ideológicas.

La preocupación es tal que la ONU, a través de sus Relatores Especiales sobre Libertad de Expresión y Opinión, junto con la Organización de Estados

torno a los beneficios y detrimentos de la imprenta. Se cuentan entre sus ventajas: la posibilidad de recopilación de obras, la superación del tiempo y el espacio, con su distribución y reparto y la contribución al conocimiento. Entre sus inconvenientes: la confusión que el exceso de obras provoca, la difusión de desatinos, falsedades, engaños en las autorías de las obras y la ayuda para desprestigiar a quienes se aborrece.

Americanos, la Organización para la Seguridad y Cooperación en Europa y la Comisión Africana de Derechos Humanos, han adoptado una Declaración Conjunta sobre *Libertad de Expresión y Noticias Falsas. Desinformación y Propaganda* el 3 de marzo de 2017. En esta Declaración se pone de manifiesto la inquietud que generan las campañas de desinformación y propaganda dirigidas a engañar a la ciudadanía. Con ellas, se interfiere en el derecho de información, a obtener un conocimiento cierto de la realidad y el derecho a la libertad de expresión. Asimismo, se pone énfasis en el peligro que este tipo de actividades pueden entrañar para la privacidad y la reputación de las personas. Y mucho más inquieta aún, lo pernicioso que pueden resultar cuando implican discriminación, hostilidad o incitación a la violencia contra determinados grupos vulnerables.

La inquietud va en aumento cuando la difusión de noticias falsas y las campañas de propaganda y desinformación provengan directamente de autoridades e instituciones públicas, a quienes corresponde fomentar un entorno favorable a la paz, la convivencia, la integración, la protección contra la discriminación y, muy especialmente, a la libertad de expresión y el derecho a la información. Al respecto, la Declaración insiste en que las restricciones a las libertades de expresión y el derecho a la información han de estar rigurosamente justificadas y que las actuaciones de los órganos públicos han de centrarse en promover un entorno de comunicaciones plurales. A su vez, y como una de las claves para abordar la desinformación y la propaganda, se recomienda a medios de comunicación y periodistas establecer sistemas efectivos de autocontrol para garantizar la veracidad de las noticias, incluyendo los derechos de corrección y réplica[3]. Otra Declaración Conjunta de los mismos órganos, adoptada en Londres el 10 de julio de 2019, recoge la desinformación entre los desafíos para la libertad de expresión en la década de los años 20 del siglo XXI[4].

Una preocupación que está también presente en el ámbito de la Unión Europea. En el año 2018 el Grupo de Expertos de Alto Nivel sobre Noticias Falsas y Desinformación de la Unión Europea, presentó su Informe sobre las *fake news* y desinformación *on line*. El grupo de expertos define la desinformación como la información falsa, inexacta o engañosa, dirigida a causar intencionadamente daños públicos y a obtener un beneficio. Con meridiana claridad se señala que la desinformación afecta negativamente al

[3] http://www.ohchr.org/SP/NewsEvents/Pages/DisplayNews.aspx?NewID=21287&LangID=E

[4] También y relacionado: "la incitación al odio; la discriminación y la violencia; el reclutamiento y la propaganda; la vigilancia arbitraria e ilegal; la interferencia respecto al uso de las tecnologías de encriptación y el anonimato, y el poder de los intermediarios en línea". Puede verse la declaración en: http://www.oas.org/es/cidh/expresion/showarticle.asp?artID=1146&lID=2

sistema democrático, sus valores y procedimientos, a la seguridad nacional, al tejido social y contribuye a socavar la confianza en la sociedad de la información y el mercado digital. Se trata, por tanto, de un problema multidimensional que cuenta en la actualidad con una infraestructura potente, destinada no sólo a producir la información defectuosa, sino también a distribuirla y amplificarla para lograr sus objetivos.

Las causas de la desinformación, las motivaciones, los actores y sus consecuencias son muy diversas, luego variadas también habrán de ser las soluciones. Entre las causas han de tenerse en cuenta, entre otras, la pérdida de credibilidad y la desconfianza en los medios de comunicación; la velocidad a la que circulan las noticias y, en consecuencia, la competencia por ser los primeros en darlas, con la consiguiente pérdida de capacidad para contrastarlas y de rigor; la falta en la red de criterios fiables para identificar fuentes de autoridad creíbles; la ausencia de instrumentos de verificación; o el recurso excesivo a titulares llamativos, con merma de precisión del hecho noticioso. Por lo que se refiere a las motivaciones, son económicas e ideológicas. Desde el punto de vista económico, la recaudación en publicidad depende de criterios cuantitativos, del número de visitas a la página que contiene la noticia, nada tiene que ver con la calidad o importancia de la misma. Desde el punto de vista ideológico, las campañas informativas pretenden una reacción que reafirme una posición ideológica o se oponga a ella. Causas y motivaciones apuntan a los distintos actores que intervienen en estos procesos de desinformación: políticos, medios de comunicación y empresas prestadoras de servicios de internet. Pero estos actores, principales beneficiarios de la manipulación informativa, no encontrarían apoyo sin una ciudadanía motivada, dispuesta a creer y reproducir aquellas comunicaciones que resultan favorables a sus ideas y que pueden debilitar posiciones ideológicamente distintas a las propias.

Partiendo de estos presupuestos, en el Informe del Grupo de Expertos se proponen algunas medidas destinadas a proporcionar soluciones a este complejo problema, tratando de evitar que se vean afectados la libertad de expresión y otros derechos fundamentales (Seijas, 2020, 1-14). Las medidas se dirigen principalmente a mejorar la transparencia, la confianza y la formación mediática e informativa de ciudadanos y profesionales. Así se contemplan iniciativas cuya finalidad es facilitar búsquedas que privilegien contenidos creíbles, que permitan identificar las fuentes de desinformación, o establecer sistemas de filtrado para acceder a información comprobada. Se trata de estimular a los medios para que se hagan con sistemas de transparencia, como el *fact-checking* u otros equipos de verificación. Sistemas y medidas que puedan potenciar la información de calidad y diluir los canales de desinformación, tratando de evitar, o al menos aminorar, el valor económico de los bulos y las noticias falsas. Sin duda, acercarse a estos objetivos requeriría mejorar la

formación y educación digital, para garantizar una lectura crítica de los contenidos a que tenemos acceso en la red. Un aprendizaje semejante podría aumentar nuestra resistencia ante las múltiples formas de desinformación; así como, favorecer la sostenibilidad y la diversidad de los medios de comunicación profesionales y serios. Finalmente, el Informe propone crear una estructura sólida de implementación y evaluación de las medidas, además de elaborar un código de buenas prácticas, específicamente orientado a afrontar y evitar la desinformación[5].

En diciembre de 2018 se produjo una Declaración Conjunta del Parlamento Europeo, el Consejo de Europa y del Comité Europeo Económico y Social, denominada "Plan de Acción contra la Desinformación". La Declaración parte de la importancia que tiene una información plural y veraz en el desarrollo de las sociedades democráticas. En esta línea de entendimiento, se considera que causa un considerable daño público la información falsa y engañosa, creada y difundida para confundir a la opinión pública y obtener beneficio económico. Las campañas de desinformación son consideradas una "batalla híbrida", que puede provenir tanto de agentes internos como externos, y que cuenta, entre sus armas específicas, con el ciberataque, las cuentas falsas o el pirateo en redes. Para luchar en esta batalla se requiere de acciones coordinadas de todos los miembros de la Unión, dispuestos a contraatacar mejorando la capacidad de sus instituciones para detectar y analizar la desinformación, fortaleciendo la coordinación y las respuestas conjuntas a los ataques previos, movilizando al sector privado frente a ellos y sensibilizando y potenciando la capacidad de respuesta de la sociedad[6].

Dentro de la adopción de medidas conjuntas para movilizar al sector privado, en ese mismo año, se había establecido un Código de Buenas Prácticas para las plataformas en línea, los anunciantes y otros agentes clave en materia de información *on line*. Quienes adopten el Código se comprometen a tomar medidas orientadas a frenar la desinformación. Entre estas medidas se encuentran el aumento de los mecanismos de transparencia y la rendición de cuentas, así como establecer el marco apropiado para realizar un seguimiento, que permita evaluar y mejorar las políticas contra la desinformación[7]. En mayo de 2021 se han elaborado unas directrices con la finalidad de reforzar el Código de Buenas Prácticas. Con ellas se trata de fomentar la colaboración de las plataformas y agentes de interés que actúan en red para acabar con la financiación de la desinformación, garantizar la integridad de los servicios

[5] Cfr. https://op.europa.eu/es/publication-detail/-/publication/6ef4df8b-4cea-11e8-be1d-01aa75ed71a1

[6] https://data.consilium.europa.eu/doc/document/ST-15431-2018-INIT/es/pdf

[7] Cfr. https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation

online, capacitar a los usuarios para comprender y denunciar la desinformación, incrementar la cobertura de verificación de datos y mejorar en la elaboración de un marco de seguimiento sólido y eficaz[8].

Todas estas iniciativas responden a la preocupación por los efectos perniciosos que la pérdida del valor de la verdad pueda tener en los derechos de la ciudadanía y sus repercusiones sociales, políticas y culturales, como consecuencia del auge de la tecnología digital. Una red libre y segura es en la actualidad un elemento clave para un eficaz disfrute de muchos derechos. Para conseguir tal propósito, es preciso que la red sea accesible, transparente y que la información que contenga sea veraz. Así pues, la veracidad de la información publicada en red constituye uno de los "bienes comunes" que debe preservarse en la era digital (Morente Parra, 2021, 212).

5. EL PODER DE LA VERDAD Y LA VERDAD DEL PODER

Algunas de las grandes preocupaciones que han mostrado los organismos internacionales, a propósito de los peligros de la desinformación en la era digital, son, de un lado, que la información defectuosa provenga de poderes públicos y, de otro lado, que las medidas para remediarlo incidan negativamente en derechos fundamentales de la ciudadanía, especialmente en la libertad de expresión y el derecho a la información.

La inquietud es lógica si se tiene en cuenta la histórica participación de algunos organismos políticos en campañas de desinformación, propaganda y, en general, la relación -más bien la mala relación- entre la verdad y el poder, la verdad y la política. De hecho, las grandes cuestiones sobre la verdad, desde perspectivas científicas y filosóficas, no permanecen ajenas al contexto del poder y de los diversos sistemas políticos que lo sustentan.

La verdad, señala Heidegger, pretende ser universal y válida para todo tiempo y lugar. Es atemporal, se mantiene en la necesidad de permanecer, pese a la multiplicidad, y carece de excepciones. De ahí que la verdad tenga "carácter del ser ideal" (Heidegger, 2004, 53). Habermas, para quien la verdad implica correspondencia con la realidad, también pone de manifiesto este carácter ideal que permite sostener la verdad. Es verdadero lo que puede ser racionalmente aceptado y resistir todos los intentos de refutación, bajo "condiciones ideales" (Habermas, 2002, 246). Circunstancias en las que la pretensión de validez universal de la verdad es conducida hacia una universal justificación (Rorty, 2007, 13, 57).

[8] Cfr. http://cde.ugr.es/index.php/union-europea/noticias-ue/1207-la-comision-presenta-unas-directrices-para-reforzar-el-codigo-de-buenas-practicas-en-materia-de-desinformacion

Pero si la verdad solo puede florecer bajo presupuestos y condiciones ideales, difícilmente podrá convertirse en una meta para la política. No en vano, y pensando en esta dificultad, Platón en la República situaba al frente del poder, en su ciudad ideal, al gobernante filósofo. Sólo los más sabios, quienes acceden al mundo de las ideas, pueden alcanzar la verdad universal. Un conocimiento de la verdad al alcance de muy pocos y que, por tanto, implica una concepción de gobierno elitista incompatible con principios básicos del sistema democrático. En este sentido, la verdad como aspiración a ser única, absoluta, dogmática e inflexible, discurriría por los parajes del totalitarismo. La verdad sería la máxima expresión del poder, propio de quien tiene el privilegio de imponer su punto de vista a los demás.

Sin embargo, en un sistema democrático resulta problemático, sino incompatible, sostener una concepción de verdad absoluta. La democracia se desenvuelve mejor entre verdades relativas y convencionales, provisionales y temporales, fluidas y flexibles. Incluso, hay quien sostiene, entre verdades maleables y útiles, alternativas e interesadas, creativas y efectivas. Dando un paso más, se podría llegar a la no verdad, si se entiende que en democracia sería posible considerar como verdad todo aquello que se haya logrado defender con éxito, que haya podido convencer al auditorio; aquello que, aun no siendo verdad, pudiera ser tomado como tal. Lo que pueda resultar verosímil, aunque no sea verdadero (Villaplana Ruiz, 2020, 27-32, 67, 70, 81). El procedimiento democrático parte de la interactuación y discusión de diversas y plurales percepciones de la realidad, a fin de generar acuerdos. Estos consensos son interpretaciones colectivas, que sugieren la construcción de relatos compartidos y reconocidos por todas las personas que intervienen en el debate. Desde estas perspectivas, no se trata, por tanto, de encontrar una verdad que ponga fin a los conflictos producidos desde interpretaciones diversas y contrapuestas, sino de hacer posible una interpretación aceptable, sin necesidad de recurrir a la violencia para imponerla. En palabras de Vattimo, la democracia implica la despedida de la verdad, un "adiós a la verdad" (Vattimo, 2010, 18-19, 29, 126, 129).

De tal manera que la verdad no se conquista tras un arduo proceso de investigación, se fabrica a partir de un acuerdo. No tendríamos hechos incontrovertibles, sino narraciones controvertidas sobre hechos. Pero, si no hay verdad, entonces tampoco habrá mentira, sino percepciones o interpretaciones diversas, puntos de vista alternativos.

En cualquier caso, la experiencia de verdad, desde un punto de vista político, estará necesariamente vinculada al poder, ya se entienda impuesta de forma totalitaria o acordada democráticamente (Vattimo, 2013, 203, 210). Cómo también estará vinculada al poder la vivencia de la mentira. Ya Platón puso de

manifiesto lo rentable que a un gobernante le resultaría hacer creer a sus gobernados que el hombre justo es feliz y el injusto infeliz (Platón, 1992, 110). Maquiavelo, en *El Príncipe*, señala la importancia que tiene para el gobernante saber disimular y fingir (Maquiavelo, 1991, 92). Max Weber apela en la política a una ética de la responsabilidad, no necesariamente coincidente con la ética de la convicción y, en este sentido, la mentira podría estar justificada (Weber, 2005, 139). Hannah Arendt, escribió que la verdad y la política nunca se llevaron demasiado bien y que nadie había incluido la veracidad entre las virtudes políticas, entre otros motivos, porque la mentira puede ser un impulso para la acción y el cambio, ambos legítimos objetivos de la política (Arendt, 1996, 239, 264, 271).

Lo cierto es que la política ha sido históricamente un terreno fértil para la mentira, en cualquiera de sus formas. La propaganda tiene un uso político, las promesas políticas se incumplen, las encuestas se "cocinan", los resultados se "maquillan". No olvidemos que el lenguaje político es básicamente un lenguaje emocional, dirigido a grandes masas de población, con grados de conocimiento muy diversos. Uno de los medios comunes que utiliza el político para tratar de convencer a su auditorio es el mitin, poco apto para la argumentación racional, con las emociones como protagonistas, proclive a las exageraciones y, entre cuyos objetivos, no está encontrar alguna verdad. En definitiva, ni a la política, ni al político le resulta ajeno un mundo de mentiras, medias verdades o posverdades (Haidar, 2018, 7).

Ahora bien, si la imposición de la verdad ha podido ser considerada como una amenaza para la democracia, puede no serlo menos la mentira recurrente. Al respecto Foucault escribió que "nada es más peligroso que un régimen político que pretende imponer la verdad" y "nada es más inconsistente que un régimen político indiferente a la verdad" (Foucault, 1991, 241). Por su parte, Arendt, advierte que el totalitarismo tiene mucho que ver con aquellos sujetos para quienes la distinción entre hechos y ficción, verdadero y falso no existe (Arendt, 2004, 289). Conviene recordar a propósito las palabras de Ricoeur: "la verdad congrega a los hombres, la mentira los dispersa y los enfrenta entre sí" (Ricoeur, 1990, 145).

Aunque no hubiera verdades absolutas, ni objetivas, o precisamente por no haberlas, el sistema democrático se sirve para construirlas de la deliberación y el consenso. Convencer es una de sus principales estrategias para su correcto funcionamiento. Cuando la verdad se convierte en algo totalmente inseguro y desconocemos donde está, tampoco sabemos dónde está la mentira. Si se pretende hacer valer como verdad cualquier cosa, acabaremos siendo escépticos respecto a ella, sin considerar nada como veraz y, por lo tanto, no habrá algo seguro sobre lo que construir (Arendt, 1996, 265-271). En un escenario tal, la

mentira deliberada encuentra el lugar apropiado para instalarse en el discurso político, y la palabra perderá gran parte de su valor para deliberar, convencer y acordar. Entonces el objetivo del discurso no será convencer, sino vencer. No importará ya la calidad del discurso, los mensajes se simplifican para alcanzar al mayor número de personas posible, la racionalidad estará dispuesta a ceder todo el espacio posible a las emociones.

Una de las consecuencias más perversas de la cadena de efectos derivados de la pérdida de valor de la verdad, es la polarización de ideas en la sociedad y la dificultad, cada vez mayor, de alcanzar acuerdos (Carvalho Feitosa Valadares, 2021, 13). En este sentido, la mentira acomodada en la sociedad y la manipulación de hechos y opiniones sobre cuestiones delicadas para la convivencia ciudadana, entrañan una forma de violencia, cuando son capaces de erosionar algunas de las bases fundamentales en que apoyar la política democrática (Rubio Núñez, 2018, 197).

La devaluación de la verdad, entraña riesgos de los que la política no está exenta y, nunca lo ha estado, de ahí la importancia para el mantenimiento en los sistemas democráticos de los contrapesos y controles, algunos estrechamente vinculados a los medios de información y comunicación. Lo peculiar en las sociedades actuales, tecnológicamente avanzadas, es, de un lado, el ritmo vertiginoso en que se difunden aquellas medio verdades o mentiras. De otro lado, y quizás más inquietante, lo irrelevante que pueden llegar a ser para la ciudadanía las mentiras y la generalización de su uso indiscriminado, aunque provengan de colectivos que antes hacían de la verdad un mérito profesional. El control de la información se ha ido dispersando, y con ello los centros de poder tradicionales, hasta el punto que más que una cruzada por la verdad, a veces, parece que la hay por conseguir el monopolio de la mentira (Márquez Guerrero, 2016).

6. UNA NUEVA CULTURA DE LA NO VERDAD: LA POSVERDAD

En otro lado de la verdad, a los términos de verosimilitud, media verdad y mentira, se ha unido, en las sociedades red, el de posverdad. La posverdad refleja una forma de aproximarse a la información, transmitirla, recibirla y asimilarla que se ha convertido en uno de los símbolos principales para reflejar y explicar algunas de las características de las sociedades del siglo XXI. Tanto es así, que se ha llegado a decir que la sociedad actual se ha instalado en una "cultura de la posverdad" (Castellanos Claramunt, 2020, 316).

Su impacto ha sido tal, que en 2016 el Diccionario de Oxford la eligió como la palabra del año, considerando que en ella "los hechos objetivos son menos determinantes que la apelación a la emoción o a las creencias personales en el modelaje de la opinión pública". El término entra en el Diccionario de la

Real Academia Española en el año 2017, con la siguiente definición: "distorsión deliberada de una realidad, que manipula creencias y emociones con el fin de influir en la opinión pública y en actitudes sociales", añadiendo que "los demagogos son maestros de la posverdad".

Esos maestros de la posverdad encuentran en la red un instrumento eficaz para practicar la demagogia, que no es sino una degeneración de la democracia, haciendo circular a gran velocidad información falsa, realidades a la carta, cargadas de contenidos emocionales, que generan un gran impacto, con un efecto perturbador, en la formación de la opinión pública. A una ciudadanía abrumada por la enorme cantidad de material circulando en la red, le resultará realmente difícil distinguir entre opinión pública y publicada, el rumor de la información y la verdad de la mentira.

De esta forma aquellos demagogos convierten la información en un mercado adulterado, en el que hacer primar la "viralidad" sobre la "veracidad" (Pauner Chulvi, 2018, 302). Parte de los contenidos en red, que pretenden ser información son, por el contrario, desinformación. Incluso, su objetivo puede tener mayor alcance que simplemente desinformar, proponiéndose confundir y deformar la opinión pública a favor de concretos fines ideológicos y políticos. En este sentido, la posverdad es considerada como una "versión posmoderna de la propaganda" (Rubio Núñez, 2018, 202). En cualquier caso, la verdad ha perdido gran parte de su trascendencia en la formación de la opinión pública y en el normal funcionamiento de los sistemas democráticos. De ahí que la posverdad no se refiere a algo que venga después de la verdad, sino a una verdad que ha perdido valor.

Esta nueva cultura de minusvaloración de la verdad viene condicionada por algunos elementos propios de la era tecnológica, que influyen decisivamente en la forma en que conocemos, en lo que conocemos y cómo lo trasladamos a la vida social y política. Sus principales ejes vertebradores son: la relativización de la verdad y la prioridad del discurso emotivo sobre el racional (Zarzalejos, 2017, 11).

6.1. La relatividad de la verdad

La negación de la existencia de verdades absolutas y su corolario, la consideración de la verdad como algo relativo, no es algo novedoso; por el contrario, forma parte de la historia del pensamiento. No obstante, han de tenerse en cuenta algunas peculiaridades propias del momento actual, estrechamente vinculadas a la información en la sociedad digital. La cantidad de información circulante, la velocidad a la que se difunde y se renueva provoca, como ya se ha señalado, que prácticamente no haya tiempo para comprobar su veracidad, ni reflexionar sobre ella y, mucho menos, para

someterla a un análisis crítico. Si a esto añadimos la pérdida de confianza en las fuentes tradicionales de transmisión de información, obtenemos como resultado una sociedad que manifiesta un gran escepticismo a distintos niveles: los medios de comunicación no proporcionan crédito, ni seguridad; no se siente tranquilidad y confianza en el funcionamiento y utilidad de la política; ni siquiera, se confía en los análisis y predicciones de la ciencia.

El rechazo de la existencia de la verdad lleva aparejado la desconfianza en los argumentos de autoridad y en las instituciones. Como consecuencia de descartar o minusvalorar todo aquello objetivo y racional, que pueda proporcionar seguridad cognoscitiva, el conocimiento científico se desacredita, al tiempo que se incrementa el protagonismo de posiciones no científicas y se expande el negacionismo científico.

Si se pierde la confianza en las instituciones, en el argumento de autoridad, en la ciencia, en los canales habituales de información y en la existencia de una verdad objetiva, solo nos queda confiar en "los nuestros". Estamos dispuestos a aceptar los relatos de las personas afines como si fueran verdad, con independencia de que los sean. Los hechos y los datos se banalizan en favor del relato de quienes piensan de forma similar, quedando a disposición del uso interesado que nos quieran trasladar y predispuestos a una adhesión incondicional. La exposición selectiva de informaciones y opiniones potencia que la comunicación funcione como una caja de resonancia en la que escuchar al otro como un eco de nosotros mismos, lo que provoca una reducción, en lugar de una expansión, de nuestras referencias y horizontes vitales (Rivero, 2016). De esta manera, se actúa una suerte de "tribalismo moral", en el que uno de los factores principales de unión es el compartir emociones, con frecuencia negativas, como la rabia, el miedo o el odio al diferente. Elementos de conexión que tienden a que se acepte todo aquello que provenga de quienes comparten nuestras ideas, aunque sea falso y seamos consciente de ello (Martínez Díaz, 2018, 448, 458). Por el contrario, estamos predispuestos a rechazar todo aquello que venga de quienes consideremos en otro "bando" ideológico, sin conceder ningún beneficio a la duda (Arias Maldonado, 2017a, 72-75).

De esta forma se produce una distorsión en la percepción de la realidad, un sesgo cognitivo, merced al cual nos protegemos de aquellas verdades que no nos gustan, nos resultan incómodas, inoportunas, contrarias a nuestros intereses o, simplemente, al modo cómo vemos o queremos ver la realidad que nos rodea. Unas verdades que ocultamos y evitamos limitando las interacciones sociales con quienes pueden proporcionarnos puntos de vista diversos, en confrontación con los nuestros. El sesgo cognitivo, como se ha señalado, nos predispone a la aceptación sin reservas de todo aquello que no implique un desafío a las ideas, creencias y sentimientos que ya tenemos. Permite reforzar

nuestra precomprensión de la realidad, lo que favorece la tendencia a evitar el esfuerzo que implicaría tener que comprobar, reflexionar y valorar detalladamente aquello que recibimos en nuestras redes de comunicación. Por tanto, en lugar de ampliar nuestro mundo cognoscente y pensante, opera restringiendo nuestro ámbito de conocimiento y nuestra capacidad de enfrentarnos a situaciones varias, disminuyendo nuestras facultades para pensar (McIntyre, 2018, 63, 82-83).

En una realidad virtual que limite las relaciones que puedan suponer complejidad deliberativa, desafío y esfuerzo intelectual, el campo de argumentación se reduce a un "me gusta", para quien nos ofrece algo coincidente o similar a nuestras ideas y creencias, o un "no me gusta", para quien nos brinda algo opuesto a ello (Dahlgren, 2012, 63) . Esta forma de proceder, proclive a la simplificación, nos sitúa en un contexto propicio para que florezcan las teorías de la conspiración, como una fácil alternativa a explicar los acontecimientos que puedan introducir cambios en nuestros entornos habituales (D´Ancona, 2019, 46-52, 107, 156) .

6.2. Prioridad de lo emocional sobre lo racional

El segundo condicionante fundamental de la llamada cultura de la posverdad es la prioridad del discurso emotivo sobre el racional y con ello el refuerzo de emociones y prejuicios (Backer, 2019, 24-26). Si alguna verdad existe no estará basada en datos objetivos, sino impregnada de sentimientos (Arias Maldonada, 2017b). Cuando los hechos se subordinan a las opiniones y la ciencia es desplazada por la ideología, las emociones toman ventaja sobre las razones, como forma de acercarse a la realidad, y como criterio guía de comportamiento. Por eso, la información en la sociedad digital da prioridad, antes que a la objetividad de hechos y datos, a la subjetividad de las percepciones y sensaciones, que permiten establecer una estrecha conexión con las emociones del destinatario (Núñez Ladevéze/Vázquez Barrio/Torrecilla Lacave, 2018, 85-86; Castells, 2010, 312). Se podrá tomar como verdad una apariencia que logre conectar con el público, con independencia de que se corresponda con la realidad (Mireles Cárdenas, 2018, 227).

El lugar privilegiado que, en la sociedad en red, ocupa la emoción frente a la razón, contribuye a reforzar el sesgo cognitivo y, también, el llamado sesgo de consenso o de confirmación. Como ya se ha señalado, el sesgo cognitivo está favorecido porque, merced a las burbujas y filtros, la información a la que normalmente accedemos contribuye a un reforzamiento de nuestras convicciones y prejuicios, al no tener accesible o rechazar aquellas propuestas y opciones que no nos resultan afines. Bajo el sesgo de confirmación, aquellas ideas fortalecidas, junto con el desconocimiento de la existencia o extensión de ideas diferentes, pueden llegar a hacernos creer que la mayoría coincide y

comparte nuestras opiniones, y generar la ilusión de la existencia de un consenso social sobre ellas (García del Muro Solans, 2019, 75-99). Como sucede con el sesgo cognitivo, el sesgo de confirmación colabora a limitar, más que a ampliar, nuestra visión del mundo.

El gran protagonismo de lo emocional -con las emociones en el centro de nuestros posicionamientos e, incluso, con el riesgo de actuar deliberadamente de forma irracional- a la vez que refuerza el sentimiento de pertenencia al grupo afín, alimenta los prejuicios hacia el diferente (Levitin, 2019, 12). La creación de relaciones que solo contribuyen a reforzar la visión propia, y nos posiciona de manera incondicional al lado de "los nuestros", dificulta considerablemente el intercambio de ideas contrarias, potenciando su rechazo. Las ideas diversas, de este modo, circulan por universos paralelos, evitando encontrarse y con tendencia a radicalizarse en sus espacios propios no compartidos. Se quedan así fuera de alcance los discursos diversos, reduciendo la posibilidad de vivir experiencias alternativas. Lejos de intentar buscar y encontrar puntos de encuentro, en estas circunstancias, la sociedad tiende a polarizarse. La polarización impide mantener flexibles los parámetros de identidad y diferencia entre individuos y grupos (Backer, 2019, 98-99), compromete el pluralismo y obstaculiza la deliberación y los acuerdos intersubjetivos (Pariser, 2017, 12-15).

7. EL LEGADO DE LA POSVERDAD

Según lo dicho, la posverdad enfrenta lo objetivo a lo subjetivo, los hechos a los sentimientos, predispone a limitar la capacidad de crítica y reflexión, conduce a simplificar las respuestas, prefiere lo radical a lo matizable, fomenta la uniformidad y polarización de las ideas, tiende a ser irreverente frente al argumento de autoridad, relativiza hechos y datos y tiende a exaltar los sentimientos para adherirse o rechazar determinadas causas (Rubio Núñez, 2010, 207-212). A los efectos de la posverdad en la democracia hay que añadir los de los algoritmos, con su enorme potencial para, desde la sombra y prácticamente sin control, colocarnos y reforzarnos en una determinada posición ideológica (O´Neil, 2017, 229-230). Características que remiten a disminuir la predisposición al diálogo y al descrédito de las instituciones (Hernández Flores, 2018, 122). En consecuencia, la cultura de la posverdad pone en jaque a la democracia y a la convivencia (Castells, 2010, 204, 306, 312) .

La cultura que se desarrolla en la red no siempre es la más propicia para que se den las condiciones de confianza y esfuerzo necesarias para que exista un debate constructivo y un diálogo. En este caso, habrá una mayor predisposición al desacuerdo y al conflicto que a la armonía y al consenso. Frente al necesario debate plural e incluyente de un sistema democrático, podemos estar abocados a una interpretación o construcción de la realidad con

tendencia a la uniformidad y a la exclusión, con el evidente riesgo de dificultar la convivencia y producir una erosión en los mecanismos de cohesión social propios de las sociedades plurales (Sánchez Cotta, 2019, 230, 234-236).

Hemos de tener en cuenta que la democracia es debate, más que elección. Es procedimiento, más que resultado. Es compromiso y lealtad, más que adhesión incondicional. Es convencer, más que imponer (Losano, 2021, 113). A apuntalar estos valores no contribuye el escenario en que prolifera la posverdad, poco proclive a la deliberación, al compromiso con lo no afín, más interesado en el resultado que en el procedimiento, poco dispuesto a convencer y a llegar a acuerdos.

Con todo, la posverdad, junto con los rumores, los bulos y las *fake news*, se han instalado en los circuitos informativos, donde se extienden, se arraigan, se aceptan y se llegan a legitimar socialmente. Una característica de la posverdad es que coloca el protagonismo en quien recibe el relato, más que en quien lo emite. Tradicionalmente, con el acento en el emisor, las mentiras han intentado ocultar la traición a la verdad, haciendo presentar el discurso por verdadero. En el relato que configura la posverdad se hace con descaro, con notoriedad, no se esconde. No es preciso disculparse al ser revelada la falta de correspondencia de lo dicho con los hechos. Ni siquiera es preciso el esfuerzo de tratar de disfrazar lo relatado como una opinión, por el contrario, se apela a otros hechos alternativos y se incide en la necesidad de adherirse a la propuesta. En definitiva, al legitimarse socialmente, los destinatarios de aquel relato de posverdad restan o no dan importancia a la debilidad o contraposición con los hechos, en que pueda sustentarse el relato. La respuesta social de la ciudadanía hacia la manipulación que la posverdad comporta, aun siendo consciente, es favorable, siempre que vaya acompañada de la carga emocional adecuada para reforzar las convicciones de su destinatario.

De ahí que posverdad ha sido calificada con una "decepcionante transparencia de la mentira" (Rodríguez Ferrándiz, 2018, 211). Mentir ha dejado de ser reprobable. La mentira se cubre de impunidad. La posverdad, en tanto mentira, es -en palabras de Victoria Camps- la peor versión de la sofística clásica que apoyada en la era de la posmodernidad, del pensamiento débil, de la sociedad líquida y la sociedad del cansancio, se despreocupa de buscar la verdad (Camps, 2017, 94-95). Nada desdeñable resulta en este contexto la irresponsabilidad que esto puede acarrear en la política (Ruíz Vicioso, 2019, 35-37). El déficit de responsabilidad está condicionado por la desvalorización de la racionalidad y una comunicación digital en el que prima el presente, lo que impide asumir las consecuencias que puedan proyectarse en el futuron (Han, 2014, 90). La legitimación de la mentira, en forma de posverdad, nos sitúa ante el riesgo de que las "realidades alternativas" se conviertan en las "alternativas

de la realidad". En la posverdad lo determinante no es ya que se trate de engañar o falsear la realidad, sino la pretensión de deslegitimarla. Se trata de defender una determinada posición aun conscientes de su falsedad (Medrán, 2017, 35; De Angelis, 2017, 38). En consecuencia, los relatos sobre diversos hechos, los hechos alternativos y la multitud de realidades que sustentan genera la contraposición de relatos, sin comunicación entre ellos, y la necesidad de imponer uno (Ruiz Vicioso, 2019, 34).

8. A MODO DE CONCLUSIÓN

El relato de la sociedad digital del Siglo XXI, en relación a las comunicaciones y su incidencia sobre las sociedades democráticas, ha mostrado luces y sombras. Será necesario mejorar la iluminación, dirigida a disipar las sombras y a fortalecer el pensamiento, para que pueda contribuir a mantener una sociedad estable y sólida, descansada y dispuesta a recuperar el camino hacia algunas verdades.

El camino ha de recorrerse a través del debate, el pluralismo, la tolerancia, el respeto, los derechos humanos y, en definitiva, del pensamiento democrático. Una mayor claridad pondrá de manifiesto que razón y emoción, objetivo y subjetivo, hechos y creencias no tienen por qué representar alternativas excluyentes como formas de acercarse a la realidad. En el ámbito social y político, dentro de un sistema democrático, las verdades absolutas dejan paso a las verdades consensuadas, históricas y sociales, basadas en el acuerdo entre ciudadanos libres e iguales. Verdades construidas a través del debate racional, sin prescindir de las emociones, de ideas y creencias plurales, e incluyentes, basadas en hechos y datos.

9. BIBLIOGRAFÍA

Alonso González, Marián (2019), "Fake News: desinformación en la era de la sociedad de la información", en: *Ámbitos, Revista Internacional de comunicación* 45, https://institucional.us.es/revistas/Ambitos/45/Mon/Fake_News_desinformacion_en_la_era_de_la_sociedad_de_la_informacion.pdf, 29-52

Arendt, Hannah (1996), "Verdad y política", en: Id., *Entre el pasado y el futuro. Ocho ejercicios sobre la reflexión política*, trad. de A. L. Poljak Zorzut, Península, Barcelona, 239-277.

— (2004), *Los orígenes del totalitarismo*, trad. de G. Solana, Taurus, Madrid.

Arias Maldonado, Manuel (2017a), "Informe sobre ciegos: genealogía de la posverdad", en: Ibánez Fanés, Jordi (ed.), *En la era de la posverdad. 14 ensayos*, Calambur, Barcelona, 65-77.

— (2017b), "Genealogía de la posverdad", en: *El País*, 30 de marzo.

Backer, Frederick de (2019), *Posverdad y fake news: propaganda y autoritarismo en el Siglo XXI*, Trabajo Fin de Máster, Facultad de Filosofía UNED, Madrid.

Bacon, Francis (2011), *La Gran Restauración (Novum Organum),* trad. de M. A. Granada, Taurus, Madrid.

Camps, Victoria (2017), "Posverdad, la nueva sofística", en: Ibáñez Fanés, Jordi (ed.), *En la era de la posverdad. 14 ensayos*, Calambur, Barcelona, 91-100.

Cardon, Dominique (2012), "El bazar y los algoritmos. Una tipología de la competencia de las métricas de la información en la web", en: Champeau, Serge y Innerarity, Daniel (comps.), *Internet y el futuro de la democracia*, Paidós, Barcelona, 211-234.

Carr, Nicholas (2011), *Superficiales. ¿Qué está haciendo Internet con nuestras mentes?*, trad. de P. Cifuentes, Taurus, Madrid.

Carvalho Feitosa Valadares, Heloisa de (2021), "Fake News e (des) informaçao: reflexões sobre o potencial da inteligencia artificial e das novas tecnologias de acelerar a erosão da democracia", en: *Teoría Jurídica Contemporánea* 6, https://revistas.ufrj.br/index.php/rjur/article/view/44812/27002, 1-29.

Castellanos Claramunt, Jorge (2020), *Participación ciudadana y buen gobierno democrático. Posibilidades y límites en la era digital*, Marcial Pons, Madrid.

Castells, Manuel (2010), *Comunicación y poder*, trad. de M. Hernández, Alianza, Madrid.

Dahlgren, Peter (2012), "Mejorar la participación: la democracia y el cambiante entorno de la web", en: Champeau, Serge / Innerarity, Daniel (comps.), *Internet y el futuro de la democracia*, Paidós, Barcelona, 45-67.

D´Ancona, Matthew (2019), *Posverdad. La nueva guerra contra la verdad y cómo combatirla*, trad. de A. Pradera Sánchez, Alianza, Madrid.

De Angelis, Carlos (2017), "Ascenso de la posverdad o cómo construir dioses a medida", en: *UNO, d+i Desarrollando ideas, Llorente & Cuenca* 27, 38-39.

Eisenstein, Elisabeth (1994), *La revolución de la imprenta en la edad moderna*, trad. de J. Bouza Álvarez, Akal, Madrid.

— (2010), *La imprenta como agente del cambio. Comunicación y transformaciones culturales en la Europa moderna temprana*, trad. de K. Bello, Fondo de Cultura Económica, México.

Foucault, Michel (1991), *Saber y verdad*, trad. de J Varela/ F. Álvarez-Uría, La Piqueta, Madrid.

Fukuyama, Francis (2003), *El fin del hombre. Consecuencias de la revolución biotecnológica*, trad. de P. Reina, Ediciones B, Madrid.

García del Muro Solans, Joan (2019), *Good bye, verdad. Una aproximación a la posverdad*, Milenio, Lleida.

García Fernández, Fernando (2007), *Ética en Internet. Manzanas y serpientes*, Rialp, Madrid.

Gleick, James (2011), *La información. Historia y realidad*, trad. de J. Rabasseda-Gascón/ T. de Lozoya, edición digital Koothrapoli.

Habermas, Jürgen (1989), *Teoría de la acción comunicativa: complementos y estudios previos*, trad. de M. Jiménez Redondo, Cátedra, Madrid.

— (2002), *Verdad y justificación*, trad. de P. Fabra / L. Díez, Trotta, Madrid.

Haidar, Julieta (2018), "Las falacias de la posverdad: desde la complejidad y la transdisciplinariedad", en: *Oxímora. Revista Internacional de ética y política* 13, 1-16.

Han, Byung-Chul (2014), *En el enjambre*, trad. de R. Gabás, Herder, Barcelona.

Heidegger, Martin (2004), *Lógica. La pregunta por la verdad*, trad. de J. A. Ciria, Alianza, Madrid.

Hernández Flores, José de Jesús (2018), "Actuación ética para orientar a la sociedad, inmersa en un laberinto de posverdad", en: Morales Campos, Estela Mercedes (coord.), *La posverdad y las noticias falsas: el uso ético de la información*, Universidad Autónoma Nacional de México, Ciudad de México, 11-132.

Hernández Pérez, Jonathan (2018), "El ecosistema de la desinformación: excesos y falsedades", en: Morales Campos, Estela Mercedes (ed.), *La posverdad y las noticias falsas: el uso ético de la información*, http://ru.iibi.unam.mx/jspui/bitstream/IIBI_UNAM/CL1005/1/09_posverdad_noticias_falsas_jonathan_hernandez.pdf, Repositorio IIBI UNAM, 203-216.

Levitin, Daniel J. (2019), *La mentira como arma. Cómo pensar críticamente en la era de la posverdad*, trad. de J. Martín Cordero, Alianza, Madrid.

Lévy, Pierre (2002), *Ciberdemocracia. Ensayo sobre filosofía política*, trad. de J. Palacio, UOC, Barcelona.

Losano, Mario G (2021), "Informática y democracia directa: ¿dirigida por quién?", en: *Derechos y Libertades* 45, 99-121.

Maquiavelo, Nicolás (1991), *El Príncipe*, trad. de M. A. Granada, Alianza, Madrid.

Márquez Guerrero, María (2016), "El trasfondo cínico de la posverdad" en: *Público* 11/12/20016, https://blogs.publico.es/dominiopublico/18745/el-trasfondo-cinico-de-la-posverdad/

Martínez Díaz, Gonzalo (2018), "La posverdad y el resquebrajamiento del orden liberal", en: *Instituto Español de Estudios Estratégicos*, Documento de Opinión 11, 441-460.

Mason, Paul (2016), *Postcapitalismo. Hacia un nuevo futro*, trad. de A. Santos Mosquera, Paidós, Barcelona.

Mathias, Paul (1998), *La ciudad de internet*, trad. de V. Pozanco, Bellaterra, Barcelona.

McIntyre, Lee (2018), *Posverdad*, trad. de L. Álvarez Canga, Cátedra, Madrid.

Medrán, Albert (2017), "En el reino de la posverdad, la irrelevancia es el castigo", en: *UNO, d+i Desarrollando ideas, Llorente & Cuenca* 27, 33-35.

Mireles Cárdenas, Celia (2018), "La posverdad a través de la prensa iberoamericana. Análisis desde las ciencias de la información documental", en: Morales Campos, Estela Mercedes (coord.), *La posverdad y las noticias falsas: el uso ético de la información*, Universidad Autónoma Nacional de México, Ciudad de México, 219-246.

Morente Parra, Vanesa (2021), "La libertad de los modernos en la sociedad digital: <El control de los datos os hará libres>", en: *Derechos y Libertades* 45, 199-231.

Morozov, Evgeny (2012), *El desengaño de internet. Los mitos de la libertad en la red*, trad. de E. G. Murillo, Destino, Barcelona.

Núñez Ladevéze, Luis / Vázquez Barrio, Tamara / Torrecilla Lacave, Teresa (2018), "La influencia de las redes sociales en la participación política en España", en: Aznar, Hugo/ Pérez Gabaldón, Marta / Alonso, Elvira / Edo, Aurora (eds.), *El derecho de acceso a los medios de comunicación. II. Participación ciudadana y de la sociedad civil*, Tirant lo Blanch, Valencia, 85-108.

O´Neil, Cathy (2017), *Armas de destrucción matemática. Cómo el big data aumenta la desigualdad y amenaza la democracia*, trad. de V. Arranz de la torre, Capitán Swing, Madrid.

Pariser, Eli (2017), *El filtro burbuja. Cómo la red decide lo que leemos y lo que pensamos*, trad. de M. Vaquero, Taurus, Barcelona.

Pauner Chulvi, Cristina (2018), "Noticias falsas y libertad de expresión e información. El control de los contenidos informativos en la red", en: *UNED, Teoría y Realidad Constitucional* 41, 297-318.

Platón (1992), *La República*, en: *Obras Completas* (Tomo IV), trad. de C. Eggers Lan, Madrid.

Rheingold, Howard (2004), *Multitudes inteligentes. Las redes sociales y las posibilidades de las tecnologías de cooperación*, trad. de M. Pino Moreno, Gedisa, Barcelona.

Richter Morales, Ulrich (2018), *El ciudadano digital. Fake news y posverdad en la era de internet*, Océano, Ciudad de México.

Ricoeur, Paul (1990), *Historia y verdad*, trad. de A. Ortiz García, Encuentro, Madrid.

Rivero, Gonzalo, "Twitter y la cámara de eco", en: Politikon, 15/02/2016.

Rodríguez Ferrándiz, Raúl (2018), *Máscaras de la mentira. El nuevo desorden de la posverdad*, Pre-Textos, Valencia.

Rorty, Richard (2007) "Universalidad y verdad", en: Id. / Habermas, Jürgen, *Sobre la verdad: ¿validez universal o justificación?*, trad. de P. Willson, Amorrortu, Buenos Aires, 9-80.

Rosnay, Joel de (1996), *El hombre simbiótico*, trad. de A. Martorell, Cátedra, Madrid.

Rubio Núñez, Rafael (2018), "Los efectos de la posverdad en la democracia", en: UNED, *Revista de Derecho Político* 103, 191-228.

Ruiz Vicioso José (2019), "Posverdad y populismo", en: *Cuadernos de Pensamiento Político* 63, 31-40.

Sánchez Cotta, Agustín (2019), "Sobre Verdad y Posverdad en sentido social", en: *Revista Internacional de comunicación* 45, 224-237.

Sartori, Giovanni (1998), *Homo videns. La sociedad teledirigida*, trad. de A. Díaz Soler, Taurus, Madrid.

Seijas, Raquel (2020), "Las soluciones europeas a la desinformación y su riesgo de impacto en los derechos fundamentales" en: *Revista de Internet, Derecho y Política* 3, 1-14.

Serna, Justo (2017), "Fake news, todo es falso salvo alguna cosa", en: Ibáñez Fanés, Jordi/ Arias Maldonado Manuel (coords.), *En la era de la posverdad. 14 ensayos*, Calambur, Barcelona, 101-116.

Vattimo, Gianni (2010), *Adiós a la verdad*, trad. de M. T. D´Meza, Gedisa, Barcelona.

— (2013), *De la realidad. Fines de la filosofía*, trad. de A. Martínez Riu, Herder, Barcelona.

Villaplana Ruiz, Javier (2020), *La posverdad a juicio. Un caso sin resolver*, Catarata, Madrid.

Weber, Max (2005), *El político y el científico*, trad. de J. Abellán García, Alianza, Madrid.

Zafra, Remedios (2017), "Redes y posverdad", en: Ibáñez Fanés, Jordi / Arias Maldonado, Manuel (coords.), *En la era de la posverdad. 14 ensayos*, Calambur, Barcelona, 181-192.

Zarzalejos, Jose Antonio (2017), "Comunicación, periodismo y fact-checking", en: *UNO, d+i Desarrollando ideas, Llorente & Cuenca* 27, 11-13.

II. Filosofía y ética de la Inteligencia artificial

CAPÍTULO VI

LAS TRASFORMACIONES DEL DERECHO EN LA ERA DE LA CIUDADANÍA DIGITAL: NUEVOS ENFOQUES Y VÍAS PARA LA DIDÁCTICA Y LA FORMACIÓN JURÍDICA

THOMAS CASADEI
CRID - Centro de Investigación Interdepartamental sobre discriminaciones y vulnerabilidad
Universidad de Módena-Reggio Emilia (Italia)
thomas.casadei@unimore.it

1. INTRODUCCIÓN

Las reflexiones que siguen recogen algunas de las consideraciones contenidas en algunos de mis estudios recientes, estructurados en forma de manual y fruto de un intenso diálogo durante años con algunos de mis colegas[1].

Me centraré en particular en cuatro aspectos: el impacto de la tecnología y la red en la experiencia jurídica (§ 1); las transformaciones en curso dentro de las profesiones jurídicas y forenses como resultado de este impacto (§ 2); la importancia de las brechas digitales en la era de la "ciudadanía digital" (§ 3); la enseñanza del derecho en la actualidad y el papel de las instituciones académicas en la definición y el apoyo de nuevos enfoques y vías (§ 4).

Se trata de poner de manifiesto, en relación con estos procesos en curso, la necesidad de una toma de conciencia generalizada que sea capaz de arraigar en las formas de entender la filosofía del derecho y sus retos actuales, así como la enseñanza de las disciplinas jurídicas y, más en general, el papel de las instituciones académicas. Estas últimas están llamadas -en consonancia con su

[1] Se trata, en particular, Casadei- Zanetti, 2020, 396-402; Casadei- Pietropaoli, 2021; Marzocco-Zullo- Casadei, 2021. De estas obras se han tomado y reelaborado, algunas partes de este trabajo. Agradezco a Raffaella Brighi, Vittorio Colomba, Francesco De Vanna, Michele Ferrazzano, Claudia Canali, Gianluigi Fioriglio, Noemi Miniscalco, Stefano Pietropaoli, Matteo Rinaldini, Silvia Salardi, Michele Saporiti, Simone Scagliarini, Iacopo Senatori, Serena Vantin, Gianfrancesco Zanetti, las conversaciones mantenidas sobre los temas que son objeto de la reflexión aquí desarrollada. Estos intercambios de conocimientos se han madurado en el ámbito de la Oficina informática DET - "Diritto Etica Tecnologie" que he contribuido a instituir, de acuerdo con su Director Gianfrancesco Zanetti, en el CRID - Centro di Ricerca Interdipartimentale su Discriminazioni e vulnerabilità de la Univerdidad de Módena y Reggio Emilia (www.crid.unimore.it). Un agradecimiento especial siento hacia Fernando H. Llano Alonso y Joaquín Garrido Martín por haberme permitido un amplio intercambio de conocimientos y opiniones con colegas españoles.

finalidad pública- a promover formas de aprendizaje continuo, incluso fuera de sus límites tradicionales.

2. EL IMPACTO DE LAS TECNOLOGÍAS INFORMÁTICAS Y DE LA RED EN LA EXPERIENCIA JURÍDICA

En la sociedad digital, el bienestar y el desarrollo humano de los individuos, como se ha señalado, "han pasado a depender significativamente de los servicios basados en los datos, la gestión del ciclo de la información y el acceso a la mercancía del conocimiento" (Faini, 2019a, XVII).

Los datos forman nuestro "yo" digital y toda actividad humana se basa en los datos (esto al menos -hay que subrayarlo- en una parte del mundo, ya que las brechas digitales entre las distintas zonas del planeta son todavía muy grandes: Ragnedda/Muschert, 2013; Peacock, 2019; Mutsvairo/Ragnedda [eds.], 2019): los procesos, a los que asistimos de forma más o menos consciente, están representados por datos procesables y capaces de impregnar todos los aspectos de la existencia de forma ubicua, integrándose en los objetos (Internet de los objetos), y ampliando las capacidades del ser humano hasta llegar a redefinir la subjetividad y la identidad personal (la "persona digital") y conseguir plasmar la configuración de nuevas subjetividades ("personalidades electrónicas", fruto de la inteligencia artificial, y robots)[2].

Los datos son "bienes jurídicos", "objeto de derechos fundamentales" pero también "objeto de contratos"; su "tratamiento", como datos personales, genera importantes problemas de consentimiento; su proliferación ha dado lugar a la aparición de un "nuevo modelo jurídico-económico", con configuraciones de poder sin precedentes en manos de las plataformas digitales[3].

Es en torno a estas cuestiones, por cierto, donde se han desarrollado algunos de los retos más relevantes que la inteligencia artificial y la robótica plantean a la filosofía del derecho[4]: reflexiones sobre la capacitación humana y sus implicaciones (Palazzani, 2015; Salardi, 2017; Borsellino [ed.] 2018; Balistreri, 2020; Salardi/Saporiti, 2020; Salardi/Saporiti/Zaganelli, 2022) pero también hay investigaciones renovadas sobre el impacto que estos procesos tienen precisamente sobre la cuestión clásica del sujeto de derecho (Casadei, 2020a, 98-100) y, más en general, a propósito de la *subjetividad* (Astone, 2020; Foglia, 2020; Maestri, 2020; Morotti, 2020), sin olvidar, por supuesto, las

[2] Sobre estos perfiles, véase el cuidado estudio de Salardi, 2020.

[3] Para un debate reciente sobre estos aspectos, véanse las contribuciones en Stanzione (ed.), 2022.

[4] Para una visión general de las cuestiones que se debaten, véanse, en una bibliografía ya extensa, los excelentes trabajos de 2020 y de Llano Alonso / Garrido Martín, 2021.

discusiones en torno al *transhumanismo*, área en la que Fernando H. ha hecho y está haciendo importantes contribuciones al debate internacional (Fernando H. Llano Alonso, 2018; 2020a; 2020b; 2021b).

Los seres humanos, los objetos inteligentes y los robots procesan e intercambian datos, se conocen y conocen a través de la información, y los efectos de estas interacciones son también múltiples a nivel social (cfr. Ferrari, 2020; Salardi, 2020); se trata de efectos que también requieren modos de aprendizaje y formación específicos y continuos. En general, es la experiencia jurídica en su conjunto, como experiencia fáctica común a cada sujeto, la que sufre profundas mutaciones.

Las consecuencias en el plano de la subjetividad son evidentes: "El yo se concibe como un sistema informativo complejo, formado por actividades, recuerdos e historias en las que se expresa nuestra autoconciencia". Desde esta perspectiva, se puede llegar a decir que "somos nuestra información" (Floridi, 2017, 78) y que la sociedad en la que vivimos es, precisamente, una *data society*: "una 'sociedad de datos', que no sólo se rige por los datos como en su advenimiento como sociedad de la información y el conocimiento, sino que está íntimamente impregnada de ellos, llegando a implicar y moldear al hombre mismo como un *data subject*" (Faini, 2019a, XVII).

En el plano de la investigación específicamente jurídico-filosófica, surgen así nuevos espacios que definen el llamado "tecno-derecho" en el que se combinan temas y problemas informáticos con la robótica jurídica (Moro/Sarra [eds.], 2017; Zaccaria, 2020) y en el que se perfila la práctica de la tecno-regulación, con cuestiones abiertas muy significativas, incluso sobre el estatus conceptual del propio derecho (Amato Mangiameli, 2017; Lettieri, 2021; Punzi, 2021; Zaccaria, 2021).

Las tranquilizadoras fronteras entre la realidad digital, a la que se accede iniciando una sesión (con el login), y la realidad analógica, a la que se accede cerrando una sesión (con el logout), "se derrumban en el concepto de realidad; los bits ocupan el lugar de los átomos; el acceso a los datos y a los servicios socava el paradigma de la propiedad de las *res* corporales; las esferas pública y privada se redefinen; la geometría del poder cambia en un mundo sin fronteras y -pero sólo aparentemente (cfr. Taplin, 2018; Pietropaoli, 2021)- de soberanos" (Faini, 2019a, XVII).

Se plantean nuevas cuestiones y problemas para el constitucionalismo (cf. De Gregorio, 2021), para la capacidad reguladora de los Estados nacionales (pero también de la Unión Europea: cf. Sánchez Bravo, 2021), surgen nuevos dilemas éticos y bioéticos (cfr. Belloso Martín, 2018; Palazzani, 2020; Llano Alonso, 2021), se avecinan nuevos retos para la justicia en sus múltiples formas

(civil, penal, internacional, en los mundos del trabajo, etc.)[5] y, más concretamente, por la justicia social (piénsese en las cuestiones relacionadas con la brecha digital (*digital divide*): Van Dijk, 2020).

Así, el derecho está llamado a reafirmar y redefinir urgentemente sus funciones: "A la luz de la connotación contemporánea de la realidad, la gobernanza de la sociedad de los datos pasa necesariamente por la gobernanza de los datos, y el derecho, encargado de regular la vida, está llamado a regular el 'diluvio de datos' que inunda la existencia contemporánea y a proteger los derechos de las personas afectadas por las diferentes declinaciones que asumen los datos, la información y el conocimiento" (Faini, 2019a, XVII; cfr. Pietropaoli, 2021).

En este contexto, los sistemas jurídicos se enfrentan a la difícil tarea de regular este "diluvio" de datos que caracteriza las experiencias de la vida contemporánea y, en particular, el derecho está llamado a regular el *big data*[6] para proteger los derechos de las personas y de la comunidad (cfr. Simoncini, 2017).

A pesar de esta exigencia, es significativo que aún no exista una regulación explícita de los *big data* en el ordenamiento jurídico europeo.

No obstante, en este sentido, son relevantes dos normativas complementarias que constituyen el marco jurídico de referencia para la libre y segura circulación de datos en la Unión Europea: el Reglamento Europeo 2016/679 de protección de datos personales y el Reglamento Europeo 2018/1807 de libre circulación de datos no personales (Faini, 2019b; cfr. Pietropaoli, 2020); sin embargo, esto confirma hasta qué punto el espacio jurídico europeo influye en la regulación de los distintos Estados nacionales (cfr. Sánchez Bravo, 2021).

En el caso del *big data*, además, la tarea a la que está llamado el derecho se hace especialmente compleja precisamente por el funcionamiento de un elemento clave del nuevo mundo: el de los *algoritmos*[7].

Estos últimos, como se ha destacado por muchos, muestran en varios aspectos un fuerte contraste con la forma tradicional de ver la realidad por parte de los juristas (Avitabile, 2017; Amato Mangiameli, 2019; Garapon/Lassègue, 2018; Casadei/Andronico [eds.], 2021; Casadei, 2022) y

[5] A título de ejemplo véanse, respectivamente, De Asís Pulido, 2021; Gómez Valdez, 2021; Bini, 2021.

[6] *Ex multis*: Holmes, 2017; Palmirani, 2020; Burri, 2021; Numerico, 2021.

[7] Véanse, a este propósito, para obtener una visión general de los problemas y cuestiones relevantes en este ámbito: Barfield, 2020; Ebers - Navas, 2020; Micklitz / Pollicino / Reichman / Simoncini / Sartor / De Gregorio (eds.) 2021; De Gregorio, 2021.

también deben ser examinados desde una perspectiva puramente didáctica y educativa, dentro de la perspectiva de la promoción de la *ciudadanía digital* (Scagliarini, 2021; Pascuzzi, 2021).

Aunque con una cierta simplificación, se puede observar que los algoritmos favorecen un método descriptivo que difiere del carácter prescriptivo propio del derecho; más ampliamente, los algoritmos se basan en metodologías deterministas, que se basan en fenómenos, circunstancias objetivas, correlaciones y probabilidades -que actúan como vectores de previsibilidad (cfr. Carleo [ed.], 2017; 2020)- con un doble efecto: si, por un lado, parecen permitir nuevas y extraordinarias potencialidades en materia de prevención, en referencia, por ejemplo, a la dimensión asistencial y al ámbito médico-sanitario -los distintos ámbitos de la e-salud (Fioriglio, 2020)- o incluso a otros perfiles relacionados con la práctica forense o, de nuevo, con la propia práctica de la libertad (cfr. Mezza, 2018; Lagioia/Sartor, 2020), por otro lado, corren el riesgo de socavar la elección individual y el libre albedrío, en los que suelen basarse los sistemas jurídicos democráticos y la normativa correspondiente.

En tal contraste, como se verá más claramente dentro de un momento, el derecho y las profesiones jurídicas se enfrentan a complejas cuestiones ético-sociales y a problemas jurídicos heterogéneos, que son tan radicales que incluso sugieren el "fin del derecho" y el "fin del jurista" (Pietropaoli, 2020).

Los ámbitos del derecho sometidos a la formalización son cada vez más numerosos.

Estos esfuerzos académicos se remontan a un campo de estudio denominado "derecho computacional", que se ha ido consolidando en la última década (cf. Ahsley, 2017; Durante, 2019). Esta expresión se refiere a "ese campo particular de la informática jurídica que se ocupa de la computabilidad del razonamiento jurídico, explorando la posibilidad de reducir las normas a un conjunto de representaciones lógicas totalmente procesables".

Se trata, como es fácil adivinar, de un enfoque que "en el plano teórico se acerca a las tesis más extremas del formalismo jurídico" (Pietropaoli, 2020a, 111) y que, en el contexto de la interacción educativa, permite, por otra parte, una importante conexión entre los perfiles teóricos y los casos prácticos: el desarrollo de los vehículos de conducción autónoma (Losano, 2017; 2019; Scagliarini [ed.], 2019; Cerini/Pisani Tedesco [eds.], 2019), en el que están invirtiendo muchas empresas automovilísticas, explica, de forma emblemática, las perspectivas actuales del derecho computacional, así como la importancia de las reflexiones sobre el mismo. Estos últimos acompañan a los nuevos proyectos tanto en el ámbito territorial como en el institucional, así como en el académico,

pero también exigen enfoques exquisitamente éticos, para examinar las consecuencias y los efectos para los espacios urbanos y para las personas (cfr., sobre este punto, Zanetti/Casadei, 2019).

Tales reflexiones se encuentran en el campo de los llamados estudios jurídicos computacionales, un área de investigación que explora experimentalmente el uso de técnicas computacionales avanzadas (inferencia *graph-based*, *machine learning*, simulación basada en agentes, computación evolutiva) para innovar los procesos de creación, comprensión y estudio del propio derecho (para una visión general, véase Lettieri, 2020: para un examen crítico: Lettieri, 2021).

Esta última, y la experiencia jurídica que con ella se relaciona, vive por tanto una fase de cambio estructural que repercute también en el ámbito profesional.

3. LAS PROFESIONES JURÍDICAS Y LAS TRANSFORMACIONES ACTUALES: ¿QUÉ TIPO DE ENSEÑANZA?

Más concretamente, en lo que respecta a las profesiones jurídicas, basta con enumerar -como ha sugerido muy oportunamente Stefano Pietropaoli- "algunas de las innovaciones tecnológicas que las han transformado rápida y radicalmente en los últimos veinte años: desde la automatización de los documentos hasta los *smart contracts*, desde las bases de datos normativas hasta la resolución de litigios en línea, desde la gestión informatizada de los procesos hasta la predicción automática y la respuesta informatizada a las preguntas jurídicas" (Pietropaoli, 2020a).

Lo que ha surgido gradualmente es una verdadera explosión de nuevas palabras: de *big data* y algoritmos, del *e-Health* y el *autonomous driving* ya se ha hablado, a estos términos podemos añadir *data privacy, net neutrality, critical data studies, neurodiritto, blockchain, smart contract, cyberwarfare*, pero también, como se verá detalladamente a continuación, una expresión che reconceptualiza las cuestiones de desigualdad, de la exclusión, de la pobreza como *digital divide*.

Se generan cuestiones radicales en el plano filosófico, pero también se trata de cuestiones que tienen un impacto preciso en el funcionamiento de quienes se encuentran en profesiones jurídicas o en ocupaciones que tienen un vínculo directo e ineludible con el derecho: piénsese en los distintos sectores de la administración pública, pero también en las implicaciones que determinan las plataformas en los distintos mundos del sistema económico y productivo (cf. Senatori, 2021; Garibaldo/Rinaldini [eds.], 2022) o, de nuevo, en el ámbito médico y sanitario (cf. Fioriglio, 2020).

Esto plantea inevitablemente cuestiones prácticas y operativas incluso para aquellos cuya función es enseñar una disciplina como el derecho, que experimenta constantemente nuevas formas y configuraciones de su propio objeto. De hecho, la experiencia jurídica está hoy en día omnipresentemente atravesada y connotada por las tecnologías de la información (para una visión general reciente: Casadei/Pietropaoli [eds.], 2021).

Veamos más en detalle estos dos aspectos, el más teórico y el operativo, por cuanto sea lícito mantener una neta distinción de planos.

En el plano filosófico, las reflexiones se refieren a las cuestiones en torno a la inteligencia artificial, es decir, lo que originalmente se llamó "cibernética" (cfr., para un *excursus*, Casadei-Pietropaoli, 2021) y que hoy se traduce en la ampliación del ámbito del "bioderecho" y en nuevas y muy relevantes controversias (Amato, 2020).

En efecto, es difícil sustraerse a la sensación de estar rodeado de términos técnicos o nombres abstractos que ya forman parte del lenguaje cotidiano. Se trata de "un nuevo tipo de palabras, que prepara y expresa una nueva era" (Pörksen, 2011, 99; cf. Andronico, 2021), o al menos una nueva forma de utilizar ciertas palabras, propia de nuestra condición histórica actual.

Si bien estas nuevas palabras son adoptadas y recordadas en muchos sectores como aspectos clave de la nueva planificación y estrategias, así como áreas de estudio indispensables, también en referencia al derecho y las disciplinas jurídicas, no faltan quienes subrayan críticamente que "más que nombrar las cosas, terminan por ocultarlas" (Andronico, 2021).

Una de estas palabras (o expresiones), que connotan un espacio constitutivamente abierto e interdependiente, ha atraído recientemente una atención particular y es la "justicia predictiva". Resulta esencial preguntarse de qué estamos hablando cuando adoptamos tal expresión. Una posible definición, entre otras que circulan, es la siguiente: "La justicia predictiva significa literalmente la justicia que prevé el futuro: es una especie de justicia anticipada. En el lenguaje común, la justicia predictiva se ha convertido en justicia predecible. Se considera que, precisamente mediante fórmulas matemáticas, se puede predecir la interpretación judicial, de acuerdo con la necesidad de seguridad jurídica, entendida precisamente no sólo como previsibilidad de la disposición legal aplicable, sino también previsibilidad del resultado judicial". (Viola, 2017)[8].

[8] Para restablecer la seguridad jurídica, se ha desarrollado un modelo que permite la interpretación de la ley, a partir de la valorización de la única disposición de la ley que se ocupa de ella expressis verbis art. 12 Preleggi. El modelo se propone como un complemento a la actividad del jurista (abogado, magistrado, notario, académico), permitiendo la identificación de

(...)

Por otro lado, hay quienes señalan que la justicia predictiva es "en verdad una etiqueta muy sintética que describe un abanico de opciones que tienen en común la aplicación de tecnologías sofisticadas tanto con fines analíticos/inductivos (descubrir patrones de decisión o pautas de comportamiento mediante el análisis y el tratamiento de datos relativos a casos y decisiones que ya han tenido lugar) como con fines prospectivos (identificar propensiones y, sobre esta base, evaluar la probabilidad de que se adopte la decisión del juez -en el caso de una solución judicial de un conflicto- o del mediador -en el caso de la activación del Adr (*Alternative dispute resolution*)- convergen en un punto que podemos definir como focal)" (Castelli/Piana, 2018, 154).

Entrar en el mundo de la justicia predictiva significa entrar en un mundo en el que la justicia se convierte en algo predecible, hasta el punto de que se puede hablar de una (especie de) justicia anticipada, y en el que la ansiada necesidad de "seguridad jurídica", entendida en términos de una previsibilidad matemática de las decisiones judiciales, puede por fin encontrar satisfacción; al fin y al cabo, un jurista de la talla de Oliver Wendell Holmes ya lo argumentó hace un siglo: el derecho no es otra cosa que la profecía del comportamiento de los tribunales (cfr. Roccaro, 2020[9]). Ahora podemos predecir realmente -digamos, anticipar- el resultado de un juicio, con certeza matemática[10].

Este tipo de resultado abre, además, numerosos interrogantes que pueden ser abordados mediante el recurso a diferentes concepciones del derecho (cfr. Casadei, 2021a; 2022b).

---

la tesis preferible (consistente con la ley), en el sentido también de la identificación de los argumentos dirimentes individuales. El modelo se presentó en la Cámara de Diputados el 30.3.2017. Cfr., para ulteriores, desarrollos: Viola (ed.) (2019).

[9] Roccaro escribe: "Parafraseando a O.W. Holmes ("The Path of the Law", en: *Harvard Law Review*, 10 (1897), 8 para el que "las profecías de lo que realmente harán los tribunales, y nada más pretencioso, es lo que entiendo por derecho", uno podría preguntar provocativamente: "las limitaciones técnicas" y la predicción de cómo "funcionarán", ¿eso y nada más es lo que se debe entender por derecho? Y quién sabe, en las aulas de Derecho podríamos acabar enseñando a rastrear una cierta regularidad en los datos producidos por las tecnologías digitales".

[10] Sin embargo, como subraya Nicola Lettieri: "La imprevisibilidad e inescrutabilidad de los modelos predictivos basados en técnicas de aprendizaje automático (*machine learning*) genera, en primer lugar, riesgos para la seguridad jurídica entendida como la posibilidad de confiar no sólo en el vigor, la duración y los efectos de las normas jurídicas, sino también en su aplicación concreta en el ámbito administrativo y judicial. La inaccesibilidad de los enunciados normativos implementados en los algoritmos y la naturaleza intrínsecamente aleatoria de las técnicas clasificatorias y predictivas evocadas anteriormente se suman así a las causas de inseguridad jurídica en las que la doctrina se ha centrado desde hace tiempo. La imposibilidad de establecer cómo y por qué se puede juzgar el riesgo de reincidencia por una herramienta de análisis predictivo (*predictive analytics*) lo ilustra de forma plástica". (Lettieri, 2021, 89; cfr. también Lettieri, 2020).

Si, por ejemplo, nos remitiésemos al *Il concetto di diritto* di Herbert Lionel Adolphus Hart encontraríamos la confirmación de que el derecho ofrece (o debería ofrecer) "razones para actuar", y no simplemente herramientas para decir primero lo que va a pasar.

La estructura constitutivamente "abierta" del lenguaje, pues, hace vana la pretensión de eliminar las incertidumbres de la interpretación mediante la positivización de sus cánones. Hart lo explica así: "Los cánones "interpretativos" no pueden eliminar estas incertidumbres, aunque pueden disminuirlas: porque estos cánones son en sí mismos normas generales para el uso del lenguaje, y hacen uso de términos generales que en sí mismos requieren interpretación. No pueden, al igual que otras normas, establecer criterios para su propia interpretación" (Hart, 2002, 149).

Por supuesto, se podría argumentar que este problema está ya superado, puesto que cuando hablamos de justicia predictiva ya no hablamos de palabras, sino de números. Si no fuera porque esta solución corre el riesgo de abrir otros problemas (Andronico, 2021).

En un plano más estrictamente operativo, los profesionales del Derecho, por su parte, ya se están enfrentando a una forma de trabajar totalmente nueva en comparación con el pasado: basta con pensar en las formas de encontrar textos legales e identificar la jurisprudencia, pero también en el apoyo que las tecnologías de la información ofrecen en la redacción de contratos y cualquier otro tipo de documento jurídicamente relevante.

La propia enseñanza del derecho, por tanto, no puede dejar de captar estas transformaciones, su alcance (teórico y de otro tipo), sus resultados (prácticos y cotidianos), su forma de cambiar, en profundidad, la propia experiencia jurídica.

Es sólo un ejemplo entre muchos otros posibles. La digitalización de la justicia conlleva dos consecuencias prácticas muy relevantes: la desmaterialización de los soportes documentales, de ahí la posibilidad de que el proceso pueda realizarse "a distancia", y la *formalización de las resoluciones*, que permite su *automatización*.

En el primer sentido, esto conduce a la realización de procedimientos judiciales a distancia, en el segundo, en cambio, a la integración parcial o total de la decisión judicial con agentes artificiales.

En referencia a Italia, a partir de los datos facilitados por el Ministerio de Justicia, se observa que el aumento de las presentaciones telemáticas de actos judiciales ha sido casi exponencial en los últimos años: de 893.265

presentaciones en 2014 a la impresionante cifra de 12.502.084 en 2020 (Costantini, 2021).

Por lo tanto, quienes están llamados a impartir cursos de derecho, así como quienes aspiran a impartir disciplinas jurídicas y económicas en los centros de enseñanza secundaria, no pueden eludir no sólo el conocimiento de estos procesos, sino también un diálogo constante con otros conocimientos y disciplinas, como la informática y la ciencia, que trabajan en torno a estas transformaciones a nivel procesal y operativo. En este contexto, una nueva enseñanza como la didáctica del derecho, introducida en Italia en 2017[11], puede desempeñar un papel esencial en la maduración de un enfoque del derecho y de la experiencia jurídica en consonancia con las transformaciones actuales.

La justicia digital requiere profesores y formadores que sepan lidiar con el impacto que la dimensión digital tiene no sólo en el derecho, sino también en la forma de enseñarlo, con referencia a los *contenidos* y, al mismo tiempo, a los *métodos técnicos*.

4. BRECHA DIGITAL: DIRECTRICES DE LA UE Y TAREAS DE LAS INSTITUCIONES

En relación con los retos mencionados anteriormente, la cuestión de las "brechas digitales" adquiere relevancia, ya que implica abordar las múltiples y a menudo superpuestas líneas de exclusión que reducen la capacidad de acceso y uso de las tecnologías, en su mayor parte debido a las desigualdades del pasado, cuyo impacto no sólo pesa sobre el perfil de la equidad sino, más generalmente, también sobre el de la eficiencia y el desarrollo económico (cf. Vantin, 2021), así como la cohesión social y, de nuevo, el acceso a las carreras profesionales y a los mundos del trabajo[12].

Se trata de una cuestión difícil de definir o enmarcar dentro de unas coordenadas inequívocas: además, no es simplemente atribuible a una lógica dicotómica (es decir, una oposición entre *haves* e *have-nots*).

Por el contrario, el fenómeno se presenta en diferentes gradaciones, según una multiplicidad de capas y niveles de interpretación.

Si al principio -es decir, desde que los periodistas de *Los Angeles Times* acuñaron el nuevo término "digital divide" (brecha digital) en 1995- la

[11] El legislador italiano ha anclado la prestación de la enseñanza del Derecho a la reforma del acceso a la función docente: esta responsabilidad formativa ha sido encomendada al área de Filosofía del Derecho (IUS/20) mediante el Decreto Ministerial nº 616 de 2017, pero perfilando un marco de objetivos mucho más amplio que el que, tradicionalmente, suponen los discursos sobre metodologías y herramientas docentes: cfr. Marzocco, 2021a.

[12] Se recogen aquí algunas consideraciones hechas en Casadei, 2022.

literatura se centró principalmente en el problema del *acceso* a las tecnologías (*physical access*), a raíz de la difusión cada vez mayor de dispositivos y fuentes de acceso a la red, así como de una serie de investigaciones publicadas a principios de la primera década del dos mil, se ha prestado mayor atención al perfil de las *competencias* en el uso de las tecnologías y, en particular, de Internet (*skills and usage*).

Más recientemente, otra línea de investigación está explorando muy adecuadamente la *cuestión de los resultados* (*outcomes*), es decir, las *desigualdades* que se derivan, como resultado, del acceso y uso desigual de la red y las tecnologías relacionadas (*material and conditional access* e *digital media use*), con respecto a varios aspectos, por ejemplo: frecuencia de uso; tiempo de permanencia en la red; diversificación de los usos; y tipo de actividades realizadas.

Desde este punto de vista, parecen especialmente pertinentes las reflexiones que se detienen en la identificación de los sujetos o grupos más expuestos al riesgo de exclusión digital, es decir, en el análisis de los diferentes tipos de brecha, como sugirió Serena Vantin[13].

Para hacer frente a los desequilibrios causados por las brechas digitales y contrarrestar las posibles discriminaciones basadas en los algoritmos o en la tecnología, la Unión Europea, desde hace unos años, ha decidido invertir en el fortalecimiento de las competencias digitales, es decir, en "tecnologías y recursos educativos abiertos", tal y como se recoge en la Resolución del Parlamento Europeo de 15 de abril de 2014[14].

Sobre los aspectos más estrictamente cognitivos y sobre el impacto de las lagunas en términos de participación democrática, el Consejo también se expresó mediante las Conclusiones de 30 de mayo de 2016 "sobre el desarrollo de la alfabetización mediática y el pensamiento crítico a través de la educación y la formación", donde se afirma que la realidad actual requiere "un acceso fácil y continuo a Internet".

En 2018, otra resolución "sobre la educación en la era digital: retos, oportunidades y lecciones que deben aprenderse para la elaboración de

[13] Vantin, 2021, en particular pp. 234-237.

[14] Este documento manifiesta una voluntad explícita de combatir la *brecha geográfica* (que afecta principalmente a los contextos rurales, montañosos o periféricos: arts. 38, 57), así como de superar la *brecha técnico-tecnológica* (arts. 11 y 22); pero prevé, en particular, medidas para superar la brecha *sociocultural y participativa*, introduciendo itinerarios de formación digital para alumnos y profesores, destinados a desarrollar una educación de calidad y universalmente accesible, así como a promover la ciudadanía activa, que precisamente pone en primer plano la dimensión digital de la ciudadanía. Sobre estos aspectos, véase Xenidis / Senden, 2020.

políticas de la UE" insistió, una vez más, en los efectos sociales de las brechas digitales, centrándose, en particular, en su impacto en el mundo del trabajo y en las vías de acceso al mismo.

Es importante señalar que el fomento del acceso digital a la educación no se traduce automáticamente en un acceso igualitario a las oportunidades de aprendizaje y que, aunque las tecnologías son cada vez más *accesibles*, la adquisición de *competencias digitales* básicas sigue siendo un obstáculo y la brecha digital persiste (cf. Sgueo, 2022).

A este respecto, los datos de Eurostat muestran que la brecha digital está lejos de ser superada, ya que el 44% de los habitantes de la Unión Europea aún carecen de las competencias digitales básicas.

En este tipo de encuestas se mencionan los grupos de personas más afectados por estos desequilibrios: adultos desempleados; personas mayores; personas con discapacidad; habitantes de zonas rurales, montañosas o periféricas; pero, sobre todo, de forma transversal, las mujeres y las niñas, con efectos que repercuten en la educación, la trayectoria profesional y la posibilidad de entrar en el mundo del emprendimiento.

El 10 de diciembre de 2019, las Conclusiones del Consejo y de los Representantes de los Gobiernos de los Estados miembros "sobre el trabajo digital de los jóvenes" reafirmaron que "es necesario reducir la brecha digital", es decir, que esta depende en gran medida del género, la edad, el nivel de educación, el grupo social y la ubicación geográfica. Para ello, recomendaron "enfoques experimentales e innovadores y nuevos modelos de cooperación para llevar a cabo actividades y servicios digitales de trabajo juvenil", aprovechando también los procesos de aprendizaje informal.

De nuevo, el 1 de diciembre de 2020, el Consejo publicó unas Conclusiones "sobre la educación digital en las sociedades del conocimiento", destinadas a poner en marcha el nuevo Plan de Acción de Educación Digital 2021-2027 de la Comisión, titulado "Repensar la educación y la formación para la era digital".

También a la luz de los efectos de la pandemia de COVID-19, durante la cual el derecho a la educación se ha comprimido en varias ocasiones, causando dificultades especialmente a los estudiantes con necesidades educativas especiales[15], pero no sólo, el documento reconoce que "la brecha digital dentro de los Estados miembros y en toda la Unión sigue siendo un desafío, ya que puede exacerbar otras desigualdades estructurales preexistentes, incluidas las socioeconómicas y de género" (Art. 19).

[15] Cf. Canali, 2020; Selva, 2020; Marzocco, 2021b, in part. 131-139; Casadei, 2022b.

Por ello, la pandemia ha replanteado con fuerza la relevancia de la relación entre brechas digitales y desigualdades, conocida en la literatura como "brecha digital de tercer nivel" (cf. Ragnedda, 2018; Ragnedda/Ruiu, 2017) y a menudo archivada, volviendo a tejer uno de los nudos en los que la democracia constitucional y social ha revelado sus profundas dificultades: la brecha digital de tercer nivel cruza "las desigualdades estructurales, es decir, las diferentes condiciones subjetivas y colectivas de acceso y posesión de ciertos recursos estratégicos, no sólo para el uso de los medios digitales *tout court*, sino para el uso de dichos medios para activar dinámicas de inclusión social (más allá de los usos recreativos, por tanto)"[16].

Si el mundo digital es una de las estructuras sociales más importantes de la sociedad contemporánea, sobre todo en tiempos de crisis como la vinculada a la propagación de COVID-19[17], las instituciones democráticas deben ser capaces de hacerle frente con eficacia mediante un enfoque multinivel que reúna perfiles técnicos y tecnológicos, perfiles de competencias, pero también perfiles económicos y sociales. Esto plantea la necesidad de una concienciación generalizada que pueda materializarse en prácticas de educación y formación permanentes y progresivas, vehiculadas, en primer lugar, por políticas públicas orientadas a una sociedad del conocimiento plenamente integradora.

La cuestión no es sólo el acceso a los espacios digitales (y democráticos), sino la capacidad de permanecer en ellos, con autonomía y capacidad de relación, practicando una *costumbre* que permita el pleno ejercicio de los derechos fundamentales, empezando por el derecho a la educación y a la formación ya no sólo para la parte inicial de la existencia, sino para la vida en su conjunto.

Por tanto, las estrategias institucionales deben orientarse hacia dos frentes: por un lado, el del apoyo y la asignación de recursos y, por otro, el de la formación siguiendo estrategias precisas de inclusión (Amato Mangiameli/Campagnoli, 2020).

5. DIDÁCTICA "SIN FRONTERAS" Y VÍAS DE ESTUDIO "HÍBRIDAS": LA PROFESIÓN DOCENTE Y EL PAPEL DE LAS INSTITUCIONES ACADÉMICAS (Y DE LOS ESTUDIOS JURÍDICOS)

Lo que hemos tratado de describir hasta ahora se refiere a varias implicaciones posibles que repercuten en la propia práctica de la enseñanza, en

[16] Selva, 2020, 466.

[17] Para un examen de las luces y sombras de la sociedad digital en el contexto pandémico, así como su impacto en la cultura jurídica, véase el completo análisis en Pérez Luño, 2020. Cfr., también, Ansuátegui Roig / Gutmann / Innerarity / La Torre (2022). También se recogen ideas útiles en Nicoletti / Lunardi (eds.), 2021.

la universidad, donde se forman los profesionales del derecho, pero también en las aulas donde trabajan quienes, a partir de los cursos de la licenciatura de derecho, van a trabajar con la dimensión de la enseñanza del derecho, también con fines profesionales.

Existe una fuerte necesidad de compenetrar los fundamentos del pensamiento jurídico, sus paradigmas más antiguos y las investigaciones sobre las funciones mismas del derecho, con la investigación y la formación continua sobre los desafíos sin precedentes que ponen en tela de juicio sus propios límites y formas, en esencia su propia identidad.

En el ámbito académico, esto debe allanar el camino hacia unos itinerarios de estudio ciertamente "híbridos", caracterizados por cursos y actividades interdisciplinares, talleres e iniciativas que pongan a los departamentos en estrecho diálogo con las órdenes profesionales, las instituciones, las organizaciones económicas y sindicales, y el amplio mundo de las asociaciones y el tercer sector, y, a través de estos métodos estudiantes con expertos y nuevas figuras profesionales comprometidas activamente -por poner un ejemplo, la del *Data Protection Officer* (DPO), ya sea en la administración pública, en el ámbito médico y académico o en el mundo económico de la empresa (cf. Angeletti, 2019; Panetta/Mauro/Sartore, 2021)- en la gestión de los efectos de las tecnologías de la información y de la inteligencia artificial, así como de los resultados de las transformaciones que afectan a los sistemas y a la vida de las personas[18].

En referencia, pues, a los retos sociales, el enfoque del *service learning*, que da fe de la concreción de la idea de una actividad formativa concebida "más allá del aula" y en el seno de la comunidad (entendida en sentido local pero también en un sentido más amplio), puede permitir no sólo concebir el territorio como espacio de participación y ciudadanía activa, además de aprendizaje (Zullo 2021b), sino también el mundo de la red como espacio de interacciones que destierren, por ejemplo, los discursos de odio y la violencia (Ansuátegui Roig, 2017; Bello/Scudieri, 2022) y las prácticas discriminatorias (Vantin, 2021).

Como se ha sugerido muy eficazmente: "Sigamos, pues, enseñando en las facultades de Derecho lo que es el abigeato, pero ayudemos también al joven alumno a encuadrar la práctica del phishing en las causas penales de nuestro ordenamiento jurídico. Insistamos en subrayar la validez invariable de la

[18] Para estas reflexiones, estoy en deuda con Stefano Pietropaoli, en el marco de un diálogo que se viene desarrollando desde hace varios años, así como con las comparaciones y debates que han madurado en los numerosos encuentros promovidos por el taller informático DET - "Derecho, Ética, Tecnología" del CRID.

definición de abigeato del *Corpus juris civilis*, pero reflexionemos también sobre la responsabilidad jurídica de un proveedor de servicios de Internet. No desechemos los voluminosos tomos de la pandectística, pero tampoco los sostengamos sobre nuestros hombros como un yugo. Apoyémonos en ellos para mirar más allá" (Pietropaoli, 2020a, 118).

Para mirar más allá -tomando prestadas las palabras que Alex Langer utilizó para describir la coexistencia civil- necesitamos "mediadores, constructores de puentes, saltadores de muros, exploradores de fronteras" (Langer, 1995, 39)[19].

Se necesitan, en este caso parafraseando a Langer, "traidores a la compacidad disciplinaria", no "tránsfugas". Es necesario dedicarse a "explorar y superar las fronteras", pero manteniendo un firme contacto con la dimensión jurídica, con las categorías y las herramientas del derecho: se trata de una actividad "decisiva para suavizar la rigidez, relativizar las fronteras, favorecer la inter-acción" (Pietropaoli, 2020a, 118) y, por tanto, para afinar las herramientas para orientarse en el mundo cambiante (y en el que cambian las dimensiones existenciales de los sujetos) o para preparar otras nuevas.

Este enfoque del derecho permite afrontar las transformaciones en curso, reclamando una forma diferente de concebir la propia *función reguladora*, cuestionando también su posibilidad de promover concretamente *principios de valor precisos*.

En cuanto al primer aspecto, el normativo, un campo que sin duda hay que cultivar es, utilizando de nuevo un ejemplo, el del *habeas data*, es decir, un instrumento dirigido a un juez para garantizar la autodeterminación informativa, la privacidad y la libertad de información de una persona: "el derecho a la libertad informativa adquiere", de este modo, "una nueva forma del tradicional derecho a la libertad personal, como derecho a controlar la información sobre la propia persona" (Frosini, 2014, 563), y reconfigura de algún modo la propia idea de ciudadanía en clave -también, pero no sólo- *digital*.

[19] Para una discusión más extensa de estos perfiles me remito a Casadei, 2021b. Langer (Sterzing 1946 - Florencia 1995) fue un político italiano, pacifista, escritor, periodista, ecologista, traductor y profesor italiano. De formación católica-socialista, más tarde exponente de la organización comunista *Lotta Continua*, dirigió también el periódico del mismo nombre. Fue uno de los fundadores del Partido Verde italiano y uno de los líderes del movimiento verde europeo, elegido diputado al Parlamento Europeo por primera vez en 1989 y reelegido en 1994. Fue promotor de numerosas iniciativas por la paz, la convivencia, los derechos humanos, contra la manipulación genética y por la defensa del medio ambiente.

En cuanto al segundo aspecto, el promocional, es -en referencia a la escuela secundaria- la idea de un derecho, por así decirlo, *sin fronteras* que puede interceptar los retos de este tipo de *ciudadanía digital*, reafirmando el sentido de una *ciudadanía activa*[20].

Esta última, inspirándose en los principios constitucionales y en la cultura de los derechos fundamentales y humanos (cfr. Zullo, 2021a; 2021b), se invoca como contrapunto a las explosiones de nacionalismo, chovinismo, racismo, fanatismo religioso, etc. -"los factores más perturbadores de la convivencia civil conocidos por el hombre" (Langer, 1995, 39)- y reafirmar su capacidad para abordar todas las dimensiones de la vida colectiva: la cultura, la economía, la vida cotidiana, así como la dimensión política o religiosa y espiritual[21].

Esta concepción del derecho permite a quienes ejercen la "profesión docente" examinar también las tensiones económicas y sociales, acercarse a quienes se encuentran en condiciones vulnerables y/o marginales (cf. Bartoli [ed.], 2019), captar los retos ecológicos, mostrar las interconexiones -demasiado a menudo oscurecidas- entre el derecho y la economía, y, más en general, esas "nuevas conexiones problemáticas", esas "instancias multifacéticas", esas "nuevas necesidades" y "relacionalidades imprevistas" que, como observó con gran clarividencia Alfonso Catania en un trabajo recientemente reeditado (Catania, 2018), constituyen el rasgo peculiar de la experiencia jurídica -y, más en general, cultural y social, económica y política- contemporánea[22].

---

[20] Para una información más detallada, véase Amato Mangiameli / Campagnoli 2020.

[21] En referencia a estas cuestiones, cabe recordar cómo, recogiendo la recomendación del Consejo de Europa, el gobierno socialista español aprobó en 2006 la introducción de una asignatura denominada "Educación para la Ciudadanía y los Derechos Humanos" (algo equivalente a la educación cívica en las escuelas italianas) para la que el filósofo del derecho y padre constituyente Gregorio Peces-Barba, junto con algunos estudiantes especialistas de la Univ. Carlos III de Madrid, elaboraron un manual: Peces-Barba, 2007. Como es sabido, la llegada de esta disciplina vino acompañada de una feroz oposición por parte de los sectores tradicionalistas y de la jerarquía eclesiástica, que se tradujo en una serie de procesos judiciales en España en todos los niveles jurisdiccionales hasta llegar al Tribunal Europeo de Derechos Humanos de Estrasburgo. La llegada al Gobierno del Partido Popular en 2012 puso fin a la larga polémica con la supresión de facto de la enseñanza a través de la Ley de Educación Pública (LOMCE) de 2013.

[22] Todo el sistema jurídico, bajo el empuje de los procesos impulsados por la globalización económica, ha perdido de hecho -como ha vuelto a señalar lúcidamente Catania- algunas de sus "características autorreferenciales (tradicionalmente investigadas por la teoría general) de plenitud, coherencia, unidad en el enfoque de la soberanía estatal": "cada vez más, el derecho se flexiona para representar instancias multifacéticas y complejas que difícilmente se pueden remontar a la organicidad de la institución estatal a la que la modernidad lo ha ligado, y por ello se hace disponible para dar voz a nuevas necesidades, a relacionalidades imprevistas, respecto de las cuales el papel de los principios de valores, positivizados en las Constituciones y

(...)

Este reto, en el día a día, permite alejarse de uno de los mayores riesgos para el profesor, que es el de -como recordaba Norberto Bobbio (1999, 36)- instalarse en la rutina, es decir, en la regularidad recurrente de las certezas establecidas.

Lo que se prefigura es un planteamiento que, de alguna manera, es bueno que se haga propio, con plena conciencia, incluso por parte de quienes trabajan a diario en los mundos del derecho y de quienes, desde la formación y los salones académicos, se preparan para entrar en ellos.

6. BIBLIOGRAFÍA

Amato, Salvatore (2020), *Biodiritto 4.0. Intelligenza artificiale e nuove tecnologie*, Giappichelli, Torino.

Amato Mangiameli, Agata C. (2017), "Tecno-regolazione e diritto. Brevi note su limiti e diritto", en: *Il diritto dell'informazione e dell'informatica*, 2, 147-167.

— (2019), "Algoritmi e big data. Dalla carta sulla robotica", en: *Rivista di filosofia del diritto*, 1, 107-124.

— / Maria Novella Campagnoli (2020), *Strategie digitali: #diritto_educazione_tecnologie*, Giappichelli, Torino.

Andronico, Alberto (2021), "Giustizia digitale e forme di vita. Qualche libro e alcune riflessioni sul nostro nuovo mondo", en: *Teoria e critica della regolazione sociale*, 2, 1-16.

Angeletti, Sauro (2019), *Data Protection Officer: una nuova professione nelle amministrazioni pubbliche?* en: *Risorse umane nella pubblica amministrazione*, 1, 32-43.

Ansuátegui Roig, Francisco Javier (2017), "Libertà di espressione, discorsi d'odio, soggetti vulnerabili: paradigmi e nuova frontiere", en: *Ars interpretandi*, 1, 29-48.

Ansuátegui Roig, Francisco Javier, Thomas, Gutmann, Daniel, Innerarity, Massimo, La Torre (2022), *Pandemia e diritti. La società civile in condizioni d'emergenza*, Edizioni Scientifiche Italiane, Napoli.

Ashley, Kevin D. (2017), *Artificial intelligence and legal analytics: new tools for law practice in the digital age*, Cambridge University Press, Cambridge.

Declaraciones de Derechos Humanos, puede constituir una referencia más significativa de lo que la rígida organización teórica deductiva podría suponer" (Catania, 2018, XX-XXI).

Astone, Antonina (2020), "La persona elettronica: verso un tertium genus di soggetto?", en: Francesco Bilotta, Fabio Raimondi (eds.), *Il soggetto di diritto. Storia ed evoluzione di un concetto nel diritto privato*, Jovene, Napoli, 253-264.

Avitabile, Luisa (2017), Il diritto davanti all'algoritmo, en: *Rivista Italiana per le Scienze Giuridiche*, 8, 315-327.

Balistreri, Maurizio (2020), *Superumani: etica e potenziamento umano*, Espress, Torino.

Barfield, Woodrow (ed.) (2020), *Cambridge Handbook on The Law of Algorithms*, Cambridge University Press, Cambridge.

Bartoli, Clelia (ed.) (2019), *Inchiesta a Ballarò. Il diritto visto dal margine*, Navarra editore, Palermo.

Bello, Barbara G., Laura Scudieri (eds.) (2022), *L'odio online: forme, prevenzione e contrasto*, Giappichelli, Torino.

Belloso Martín, Nuria (2018), "La necesaria presencia de la ética en la robótica: la roboética y su incidencia en los derechos humanos", en: *Cadernos do Programa de Pos-Graduaçao em Direito*, 2, 81-121.

Bini, Stefano (2021), "Algoritmos y abogacía digital: reflexiones sobre el cambio de paradigma en el trabajo del abogado contemporáneo", en: Fernando H., Llano Alonso, Joaquim, Garrido Martín (eds.), *Inteligencia artificial y derecho. El jurista ante los retos de la era digital*, Thomson Reuters Aranzadi, Navarra, 51-65.

Bobbio, Norberto (1999), "Il mestiere di vivere, il mestiere di insegnare, il mestiere di scrivere, colloquio con P. Polito", en: *Nuova Antologia*, 2211, 5-47.

Borsellino, Patrizia (ed.) (2018), "Il *potenziamento umano* come ultima frontiera della biomedicina: considerazioni critiche in prospettiva etica e giuridica", en: *Rivista di filosofia del diritto*, 2, 215-272 (con contributi di P. Borsellino, L. Palazzani, S. Salardi, F. Pizzetti).

Burri, Mira (2021), *Big data and global trade law*, Cambridge University Press, Cambridge.

Campione, Roger (2020), *La plausibilidad del derecho en la era de la inteligencia artificial: filosofía carbónica y filosofía silícica del derecho*, Dykinson, Madrid.

Canali, Claudia, "Gli effetti del digital divide durante la pandemia da COVID-19", en: Porro, Carlo A., Paolo Faloni (eds.), *Emergenza COVID-19: impatto e prospettive*, cit., 69-84.

Carleo, Alessandra (ed.), (2017), *Calcolabilità giuridica*, il Mulino, Bologna.

— (2020), *Decisione robotica*, il Mulino, Bologna.

Casadei, Thomas (2021a), "Il diritto in azione: significati, funzioni, pratiche", en: Marzocco, Valeria, Silvia Zullo, Thomas Casadei, *La didattica del diritto. Metodi, strumenti, prospettive*, Pacini, Pisa, 89-120.

— (2021b), "L'impatto delle tecnologie informatiche e della rete sull'esperienza sociale e giuridica", en: Marzocco, Valeria, Silvia Zullo, Thomas Casadei, *La didattica del diritto. Metodi, strumenti, prospettive*, 156-173.

— (2021c), "Una didattica "senza frontiere"? Le trasformazioni del diritto e il mestiere dell'insegnante", en: Blengino, Claudia P., Claudio Sarzotti (eds.), *Quale formazione per quale giurista? In-segnare il diritto nella prospettiva socio-giuridica*, Quaderni del Dipartimento di Giurisprudenza, Torino, 77-93.

— (2022a), "Istituzioni e algoritmi: tra strategie funzionali ed «effetti collaterali»", en: U. Salinitro (ed.), *Smart. La persona e l'infosfera*, Pacini, Pisa, in corso di pubblicazione.

— (2022b), "«Una questione di accesso»? Democrazia e nuove tecnologie. Il caso dell'istruzione", en: Salardi, Silvia, Saporiti Michele, Vetis Zaganelli Margareth (eds./organização de), *Diritti umani e tecnologie morali: una prospettiva comparata tra Italia e Brasile/Direitos Humanos e tecnologias morais: uma perspectiva comparada entre Itália e Brasil*, Giappichelli, Torino, in corso di pubblicazione.

Casadei, Thomas, Gianfrancesco Zanetti (2020), *Manuale di Filosofia del diritto. Figure, categorie, contesti*, Giappichelli, Torino.

Casadei, Thomas, Alberto Andronico (eds.) (2021), "Algoritmi ed esperienza giuridica", en: *Ars interpretandi*, 1, 7-11.

Casadei, Thomas, Stefano Pietropaoli (eds.) (2021), *Diritto e tecnologie informatiche. Questioni di informatica giuridica, prospettive istituzionali e sfide sociali*, Wolters Kluwer, Padova.

Castelli, Claudio, Daniela Piana (2018), "Giustizia predittiva. La qualità della giustizia in due tempi", en: *Questione giustizia*, 4, 153-165.

Catania, Alfonso (2018), "Trasformazioni del diritto in un mondo globale", en: Id., *Effettività e modelli normativi. Studi di filosofia del diritto*, ed. V. Giordano, Giappichelli, Torino, 165-182.

Cerini, Diana, Pisani Tedesco, Andrea (eds.), (2019), *Smart mobility, smart cars e intelligenza artificiale: responsabilità e prospettive*, Giappichelli, Torino.

Costantini, Federico (2021), "Giustizia elettronica e digitalizzazione giudiziale: contesto europeo ed esperienza italiana", en: Casadei, Thomas, Stefano Pietropaoli, *Diritto e tecnologie informatiche. Questioni di informatica giuridica, prospettive istituzionali e sfide sociali*, cit., 118-132.

De Asís Pulido, Miguel (2021), "Derecho al debido proceso e inteligencia artificial", en: F.H. Llano Alonso, Joaquin, Garrido Martín (eds.), *Inteligencia artificial y derecho. El jurista ante los retos de la era digital*, 67-89.

De Gregorio, Giovanni (2021), *Digital Constitutionalism in Europe. Reframing Rights and Powers in the Algorithmic Society*, Cambridge University Press, Cambridge.

Durante, Massimo (2019), *Potere computazionale: l'impatto delle ICT su diritto, società e sapere*, Milano, Meltemi.

Ebers, Martin, Navas, Susana (eds.) (2020), *Algorithms and Law*, Cambridge University Press, Cambridge.

Faini Fernanda (2019a), Data society. *Governo dei dati e tutela dei diritti nell'era digitale*, Giuffrè, Milano.

— (2019b), "Big data, algoritmi e diritto", en: *DPCE on line,* 3, 1869-1882: http://www.dpceonline.it/index.php/dpceonline/article/view/785/726.

Ferrari, Vincenzo (2020), "Note socio-giuridiche introduttive per una discussione su diritto, intelligenza artificiale e big data", en: *Sociologia del diritto*, 3, 9-32.

Fioriglio, Gianluigi (2020), *Informatica medica e diritto. Un'introduzione*, Mucchi editore, Modena.

Floridi Luciano (2017), *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, Raffaello Cortina Editore, Milano.

Foglia, Massimo (2020), "L'identità personale nell'era della comunicazione digitale", en: Francesco, Bilotta, Fabio Raimondi (eds.), *Il soggetto di diritto*, 265-276.

Frosini, Tommaso Edoardo (2014), "Google e il diritto all'oblio preso sul serio", en: *Diritto dell'Informazione e dell'Informatica*, 4/5, 563-567.

Garapon, Antoine, Lassègue, Jean (2018), *Justice digitale. Révolution graphique et rupture anthropologique*, PUF, Paris.

Garibaldo, Francesco, Rinaldini, Matteo (eds.) (2022), *Il lavoro operaio digitalizzato. Inchiesta nell'industria metalmeccanica bolognese*, il Mulino, Bologna.

Gómez Valdez, Manuel A. (2021), "Si tan sólo tuviera en cerebro. La inteligencia artificial frente al derecho internacional", en: Fernando H. Llano Alonso, Joaquim, Garrido Martín (eds.), *Inteligencia artificial y derecho. El jurista ante los retos de la era digital*, 67-89.

Hart, H.L.A. (2002), *Il concetto di diritto* (1961), Einaudi, Torino.

Holmes, Dawn E. (2017), *Big data: a very short introduction*, Oxford University Press, Oxford.

Langer Alexander (1995), *La scelta della convivenza*, edizioni e/o, Roma.

Lettieri, Nicola (2020), "Law in The Turing's Cathedral. Notes on the Algorithmic Future of Legal Research", en: W. Barfield (ed.), *Cambridge Handbook on The Law Of Algorithms*, Cambridge University Press, Cambridge, 2020, 32-95.

— (2021), *Antigone e gli algoritmi. Appunti per un approccio giusfilosofico*, Mucchi, Modena.

Lagioia, Francesca, Giovanni, Sartor (2020), "Profilazione e decisione algoritmica: dal mercato alla sfera pubblica", en: *Federalismi.it*, 11, 85-110.

Llano Alonso, Fernando H. (2018), *Homo Excelsior. Los límites ético-jurídicos del transhumanismo*, Tirant lo Blanch, Valencia.

— (2020a), "Transhumanismo, vulnerabilidad y dignidad humana", en: Álvaro A. Sánchez Bravo (ed.), *Derecho, inteligencia artificial y nuevos entornos digitales*, Punto Rojo, Sevilla, 45-74.

— (2020b), "Transhumanismo y Nuevas Tecnologías: un nuevo paradigma en el ámbito de la revolución 4.0", en: *Tiempo de paz*, 138, 28-35.

— (2021a), "De máquinas y hombres. Tres questiones ético-jurídicas sobre la inteligencia artificial", en: Fernando H. Llano Alonso, Joaquín, Garrido Martín (eds.), *Inteligencia artificial y derecho. El jurista ante los retos de la era digital*, 201-234.

— (2021b), "Transhumanismo e identidad humana ante el reto de la revolución 4.0 de la inteligencia artificial", en: Luis Manuel, Lloredo Alix, Alessandro, Somma (eds.), *Scritti in onore di Mario G. Losano: Dalla filosofia del diritto alla comparazione giuridica*, Academia University Press, Torino, 325-370.

— (2021c), "L'etica dell'intelligenza artificiale nel quadro giuridico dell'Unione europea", en: *Ragione pratica* 2/2021, 327-348.

Losano, Mario G. (2017), "Il progetto di legge tedesco sull'auto a guida automatizzata. Appendice: Il progetto di legge e le relazioni illustrative", en: *Il diritto dell'informazione e dell'informatica*, 1 ss.

— (2019), *"Verso l'auto a guida autonoma in Italia"*, en: *Il diritto dell'informazione e dell'informatica*, 2, 423-441.

Maestri, Enrico (2020), "La persona digitale tra *habeas corpus* e *habeas data*", en: Francesco, Bilotta, Fabio, Raimondi (eds.), *Il soggetto di diritto,* Jovene, Napoli, 277-290.

Marzocco, Valeria (2021a), "Insegnare il diritto. Il quadro delle fonti normative e la sua evoluzione", en: Marzocco, Valeria, Silvia Zullo, Thomas Casadei (2021), *La didattica del diritto. Metodi, strumenti e prospettiva*, Pacini, Pisa, 1-48.

— (2021b), "Didattica del diritto e formazione giuridica: una mappa dei problemi", en: Marzocco, Valeria, Silvia Zullo, Thomas Casadei (2021), *La didattica del diritto*, *Metodi, strumenti e prospettiva*, Pacini, Pisa, 125-139.

Mezza, Michele (2018), *Algoritmi di libertà. La potenza del calcolo tra dominio e conflitto*, Donzelli, Roma.

Micklitz, Hans-W., Pollicino, Oreste, Reichman, Amnon, Simoncini, Andrea, Sartor, Giovanni, De Gregorio, Giovanni (eds.) (2021), *Constitutional Challenges in the Algorithmic Society*, Cambridge University Press, Cambridge.

Moro, Paolo, Claudio, Sarra (eds.) (2017), *Tecnodiritto: temi e problemi di informatica e robotica giuridica*, Franco Angeli, Milano.

Morotti Emanuela (2020), "Una soggettività a geometrie variabili per lo statuto giuridico dei robot", en: Fabio Bilotta, Fabio, Raimondi (eds.), *Il soggetto di diritto*, cit., 291-306.

Mutsvairo, Bruce, Ragnedda Massimo (eds.) (2019), *Mapping the digital divide in Africa: a mediated analysis*, Amsterdam University Press, Amsterdam.

Nicoletti, Michele, Lunardini, Marianna (eds.) (2021), *Pandemia e diritti umani. Fra tutele ed emergenza*, Donzelli, Roma.

Numerico, Teresa (2021), *Big data e algoritmi: prospettive critiche*, Carocci, Roma.

Palazzani, Laura (2015), *Il potenziamento umano: tecnoscienza, etica e diritto*, Giappichelli, Torino.

— (2020), *Tecnologie dell'informazione e intelligenza artificiale. Sfide etiche al diritto*, Studium, Roma.

Panetta, Rocco, Tommaso Mauro, Federico Sartore (2021), *Il data protection officer tra regole e prassi*, prefazione di Guido Scorza, Giuffrè Francis Lefebvre, Milano.

Pascuzzi, Giovanni (2021), *La cittadinanza digitale: competenze, diritti e regole per vivere in rete*, il Mulino, Bologna.

Peacock, Anne (2019), *Human rights and the digital divide*, Routledge, London-New York.

Peces-Barba Gregorio (2007), *Educación para la ciudadanía y derechos humanos*, con la colaboracion de E. Fernandez, R. de Asís y F.J. Ansuátegui Roig, Espasa, Madrid.

Pérez Luño, Antonio Enrique (2020), "La inteligencia artificial en tiempo de pandemia", en: Fernando H., Llano Alonso, Joaquín, Garrido Martín (eds.), *Inteligencia artificial y derecho. El jurista ante los retos de la era digital*, 33-50.

Pietropaoli, Stefano (2020a), "Fine del diritto? L'intelligenza artificiale e il futuro del giurista", en: Stefano, Dorigo (ed.), *Il ragionamento giuridico nell'era dell'intelligenza artificiale*, Pacini, Pisa, 107-118.

— (2020b), "Habeas data. I diritti umani alla prova dei big data", en: Sebastiano Faro, Tommaso Edoardo Frosini, Ginevra Peruginelli (eds.), *Dati e algoritmi. Diritto e diritti nella società globale*, il Mulino, Bologna, 97-111.

— (2021a), "Da cittadino a user. Capitalismo, democrazia e rivoluzione digitale", en: Anna Cavaliere, Geminello, Preterossi (eds.), *Capitalismo senza diritti?*, Mimesis, Milano-Udine, 31-41.

— (2021b), "Sfide attuali di un futuro prossimo: tre questioni", en: *Pietropaoli, Stefano, Faini Fernanda, Scienza giuridica e tecnologie informatiche*, n.e., Giappichelli, Torino, 517-542.

Pörksen, Uwe (2011), *Parole di plastica. La neolingua di una dittatura internazionale*, Textus, L'Aquila.

Punzi, Antonio (2021), "Difettività e giustizia aumentata. L'esperienza giuridica e la sfida dell'umanesimo digitale", en: *Ars Interpretandi*, 1, 113-128.

Ragnedda, Massimo (2018), *The third digital divide: A weberian approach to digital inequalities*, Routledge, New York-London.

Ragnedda, Massimo, Glen W., Muschert (2013), *The Digital Divide. The internet and social inequality in international perspective*, Routledge, London and New York.

Ragnedda, Massimo, Maria Laura Ruiu (2017), "Social capital and the three levels of digital divide", en: Massimo, Ragnedda, Glen W. Muschert (eds.), *Theorizing Digital Divides*, Routledge, New York-London, 21-34.

Roccaro, David (2020), "Tecnologie normative: verso un diritto avvolto dal digitale", en: *Jura Gentium*, 2, 114-123.

Salardi, Silvia (2017), Destined to be super human? Moral Bioenhancement and its legal viability, en: *BioLaw Journal - Rivista di Biodiritto*, 3, 87-101.

— (2020), "Robótica e inteligencia artificial: retos para el Derecho", en: *Derechos y Libertades*, 42, 203-232.

Salardi, Silvia, Saporiti Michele (2020), *Le tecnologie "morali" emergenti e le sfide etico-giuridiche delle nuove soggettività*, Giappichelli, Torino.

Salardi, Silvia, Saporiti Michele, Vetis Zaganelli Margareth (eds.) (2022), *Diritti umani e tecnologie morali: una prospettiva comparata tra Italia e Brasile/Direitos Humanos e tecnologias morais: uma perspectiva comparada entre Itália e Brasil*.

Sánchez Bravo, Álvaro (2021) "Inteligencia artificial, control y nuevos marcos normativos en la Unión europea", en: Fernando H., Llano Alonso, Joaquín, Garrido Martín (eds.), *Inteligencia artificial y derecho. El jurista ante los retos de la era digital*, 307-330.

Scagliarini, Simone (2021), *I diritti costituzionali nell'era di internet: cittadinanza digitale, accesso alla rete e* net neutrality, en: Casadei, Thomas, Stefano Pietropaoli (eds.), *Diritto e tecnologie informatiche*, 3-15.

— (ed.) (2019), *Smart roads e driverless cars: tra diritto, tecnologie, etica pubblica*, Giappichelli, Torino.

Selva, Diana (2020), "Divari digitali e disuguaglianze in Italia prima e durante il COVID-19", en: *Culture e Studi del Sociale*, 2, 463-483.

Senatori, Iacopo (2021), " 'Remoto' e 'multilocale': l'impatto della trasformazione digitale nei mondi del lavoro", en: Casadei, Thomas, Pietropaoli, Stefano (eds.), *Diritto e tecnologie informatiche. Questioni di informatica giuridica, prospettive istituzionali e sfide sociali*, 91-104.

Sgueo, Gianluca (2022), *Il divario. I servizi pubblici digitali tra aspettative e realtà*, Egea, Milano.

Simoncini, Andrea (2017), "Sovranità e potere nell'era digitale", en: Frosini, Tommaso E, Oreste Pollicino, Ernesto Apa, Marco Bassini (eds.), *Diritti e libertà in internet*, Le Monnier università-Mondadori education, Firenze, 19ss.

Stanzione, Pasquale (2022), *I "poteri privati" delle piattaforme e le nuove frontiere della privacy*, Giappichelli, Torino.

Taplin, Jonathan (2018), *Amazon, Google, Facebook: i nuovi sovrani del nostro tempo* (2017), Macro, Cesena (FC).

Van Dijk, Jan (2020), *The digital divide*, Polity, Cambridge.

Vantin, Serena (2021), "Digital divide. Discriminazioni e vulnerabilità nell'epoca della rete globale", en: Casadei, Thomas, Stefano Pietropaoli (eds.), *Diritto e tecnologie informatiche*, 233-245.

Viola, Luigi (2017), *Interpretazione della legge con modelli matematici. Processo, a.d.r., giustizia predittiva*, Centro Studi Diritto Avanzato, Milano.

— (ed.) (2019), "*Giustizia predittiva e interpretazione della legge con modelli matematici*, Atti del convegno tenutosi presso l'Istituto dell'Enciclopedia Italiana Treccani, Centro Studi Diritto Avanzato, Milano.

Xenidis, Raphaële, Senden, Linda (2020), "EU non-discrimination law in the era of artificial intelligence. Mapping the challenges of algorithmic discrimination", en: Ulf, Bernitz, Xavier, Groussot, Jaan, Paju, Sybe A. De Vries (eds.), *General principles of EU law and the EU digital order*, Kluwer Law International, Alphen aan den Rijn, 151-182.

Zaccaria, Giuseppe (2020), "Figure del giudicare: calcolabilità, precedente, decisione robotica", en: *Rivista di diritto civile*, 66 (2), 277-294.

— (2021), "Mutazioni del diritto: innovazione tecnologica e applicazioni predittive", en: *Ars Interpretandi*, 1, 29-52.

Zanetti, Gianfrancesco, Thomas Casadei (2019), "Tra dilemmi etici e potenzialità concrete: le sfide dell'*autonomous driving*", en: Simone Scagliarini (a cura di), *Smart roads e driverless cars: tra diritto, tecnologie, etica pubblica*, Giappichelli, Torino, 41-54.

Zullo, Silvia (2021a), "La didattica del diritto tra teorie dell'apprendimento, orientamenti pedagogici e strategie per l'insegnamento scolastico", en: Marzocco, Valeria, Silvia, Zullo, Thomas, Casadei, *La didattica del diritto*, 40-87.

— (2021b), "Lo spazio delle scienze umane e sociali nel laboratorio didattico-giuridico", en: Marzocco, Valeria, Silvia, Zullo, Thomas, Casadei, *La didattica del diritto*, 139-156.

CAPÍTULO VII

INTELIGENCIA (ARTIFICIAL) Y AUTOMATISMO. ANATOMÍA DE UN CONFLICTO

JOAQUÍN GARRIDO MARTÍN
Universidad de Sevilla
jgmartin@us.es

Ya,
grande pensamiento mío,
que estamos solos los dos,
hablemos claro yo y vos,
pues solo de vos confío.
Mi albedrío, ¿es albedrío
libre o esclavo?
-Calderón, "La hija del aire" (1664)

1. CIENCIAS COGNITIVAS E INTELIGENCIA ARTIFICIAL

Las ciencias cognitivas y la inteligencia artificial son disciplinas hermanadas desde sus momentos fundacionales en tiempos de posguerra. Eran tiempos de Hal 9000, el ser maquinal imaginado por Kubrick en 2001. No hemos llegado a este tipo Inteligencia artificial "fuerte" o general (Goertzel/Pennachin, 2009), como quieren hoy llamarla los programadores que ven próximo el umbral de la máquina superinteligente o el momento de la "singularidad", pero lo que sí se ha desarrollado desde entonces en una línea que da comienzo en aquel momento y que se perpetúa hasta nuestros días es una determinada idea de la mente que responde a los parámetros mecanicistas de la máquina computacional. La idea de John McCarthy en los albores fundacionales de la inteligencia artificial venía a ser esta misma; en la propuesta que entregaba para el "Proyecto de investigación sobre inteligencia artificial" (1955) el gran matemático explicaba: "Cada aspecto del aprendizaje o cualquier otra característica de la inteligencia puede, en principio, describirse con tanta precisión que se puede hacer que una máquina lo simule" (McCarthy, 2006, 12). El silogismo es relativamente sencillo: si averiguamos cómo funciona la mente, una máquina será capaz de replicarla. De fondo, el prejuicio del automatismo, que no distingue entre organismo y mecanismo, entre la inteligencia de la conciencia y el automatismo de la máquina.

Como explicó Bernard Stiegler (2002)[1], las propias historias paralelas de la tecnología informática y la inteligencia artificial han dado lugar a las ciencias

[1] Vid. además Gardner, 1985 y Dupuy, 2009.

cognitivas tal y como las conocemos hoy, es decir como el estudio de la cognición humana visto a la luz del modelo del ordenador. "La consecuencia de esto es que el concepto de máquina abstracta exportado desde el dominio matemático en el que es elaborado hacia un contexto tecnológico donde es reemplazado constituye un vector heurístico original y fecundo para la comprensión de los fenómenos cognitivos en general" (Stiegler, 2002, 244). El esquema lo encontramos ya en el famoso "Grupo Cibernético" de mediados del siglo pasado, reunido en torno a las famosas "Macy Conferences" (1946-53) que habían de poner en diálogo a grandes personalidades de la ciencia. Matemáticos, lógicos, ingenieros, fisiólogos, neurofisiólogos, psicólogos, antropólogos y economistas se reunían para construir de este modo transversal una ciencia general sobre el funcionamiento de la mente humana (a esto llamamos "ciencias cognitivas"). La idea fundamental que vertebró aquellos encuentros consistía en concebir el pensar como una forma de computación que no se cifra en la sola manipulación de símbolos a que aplicar reglas, sino que se relaciona más bien con esa forma de computar propia de las máquinas que técnicamente se denominan algoritmos (Heims, 1991). El pensamiento entraba así en el ámbito de lo mecánico. Con ello pensaron sería finalmente posible construir una teoría científica de la mente, dando solución al viejo problema filosófico de la relación entre la mente y la materia. En su ambición de encontrar el punto de unión entre el mundo del sentido con el de las leyes físicas no fueron desde luego originales; la oposición clásica entre materialistas y dualistas había tocado ante todo este tema sensible del alma y su carácter in/material: filósofos de uno y otro bando discutían las razones últimas de la inteligencia humana, hipotecados todos por unos presupuestos metafísicos bien conocidos; los materialistas temían un dualismo que condujera directamente a un ámbito de trascendencia ligada al fenómeno religioso, que rechazaban; los dualistas veían por su parte en el materialismo una amenaza a la libertad de arbitrio, determinada la voluntad por las solas pautas de materia y movimiento.

Estos nuevos científicos creían actuar en cambio liberados de ataduras apriorísticas. Para lograr este objetivo se volvía necesario construir un modelo de mente, pues solo se habría entendido el fenómeno de la inteligencia humana si este era susceptible de replicarse. Se ambicionaba así la creación de un cerebro que acompañara todas las características de la mente humana. Recordemos una de las obras clásicas de la época de la llamada cibernética, que llevaba el título *Diseño para un cerebro*, de W. Ross Ashby (1948); en estos términos elocuentes se expresaba el célebre cibernetista: "la fabricación de un cerebro sintético requiere ahora algo más que tiempo y trabajo. …una máquina así podría usarse en el campo de la investigación... Para explorar regiones de complejidad y sutileza intelectual más allá de los poderes humanos… ¿cómo

acabará? Sugiero que la forma más sencilla de averiguarlo es hacer la cosa y ver" (1948, 382-83).

2. ANALOGÍAS CONTEMPORÁNEAS CEREBRO-MÁQUINA

Si trasladamos el esquema a nuestro presente, advertimos la presencia de este mismo modelo en la aproximación contemporánea al problema de la inteligencia, que es así vista a la luz de su hermana la *Artificial Intelligence*. En un autor de reconocido prestigio académico y éxito editorial como Steven Pinker, generosamente traducido a diversas lenguas incluida la nuestra, la presencia de esta idea de la "mente mecánica" se nos muestra de forma transparente en su celebrado libro *How the Mind Works*: "La mente es lo que hace el cerebro; concretamente, el cerebro procesa la información, y el pensamiento es un tipo de cálculo" (Pinker, 1997, 21). Para el eminente profesor de ciencias cognitivas en Harvard la mente es vista como el resultado del proceso mecánico desarrollado en el cerebro. Y la mecanización de la mente nos viene revelada por el funcionamiento cerebral, un cerebro que viene siendo objeto de investigación masiva en los últimos tiempos -a los 90 se los denominó "la década del cerebro", y a este siglo XXI se le ha denominado el siglo del cerebro (Vidal, 2011, 358)[2].

En efecto, la época digital ha traído el mapeo del cerebro. Modelado según las más desarrolladas simulaciones de la tecnología digital, el cerebro contemporáneo deviene en cerebro digital (Dumit, 2004; Alac, 2011): al estudiarse a través de tecnologías virtuales basadas en visualizaciones digitales, el cerebro adopta las características propias de las tecnologías que lo representan, volviéndose así espejo de la tecnología que lo visualiza[3]: "Los visuales científicos digitales son campos de interacción, ya que hay que entenderlos con respecto a cómo se trabaja con ellos y se experimentan. En otras palabras, su carácter no es necesariamente representacional, sino que se refiere a la participación de sus lectores/escritores"[4]. Este cerebro es visto, además, como una especie de ordenador neuronal que en sus complejos cálculos sigue las pautas regladas de la computación mecánica. Y así, adopta el modelo de la

[2] La nuestra es la época, se ha dicho, de la "ideología cerebral", donde nuestra "personhood" ha dado paso a la nueva "brainhood", que viene a esenciar el sujeto moderno: un "cerebral subject": Vidal, 2009. Para una perspectiva histórica vid. Hanger, 1997.

[3] "Las imágenes de fMRI (resonancia magnética funcional) no son signos icónicos en términos de la idea ingenua de similitud, sino que generan significado al apoyarse en una variedad de estructuras semióticas que funcionan como su "infraestructura para ver". La figura publicada de la resonancia magnética no "revela" directamente a un ojo pasivo el cerebro y sus procesos; en cambio, se basa en una variedad de signos que indican lo que la figura muestra, ya que apelan al conocimiento cultural y al compromiso experiencial de los espectadores": Alac, 2011, 37.

[4] Ibid.

máquina. La moderna disciplina científica de la neurociencia computacional es el signo visible de este modo de ver las cosas, tenida por una "especialización dentro de la neurociencia teórica que emplea ordenadores para estimular modelos" (Trappenberg, 2010, 2)[5]. El modelo lo vemos hoy emblemáticamente en los famosos proyectos del cerebro humano estadounidense y europeo (los conocidos "Brain Iniciative" y "Human Brain Project"). Su misión, se afirma, es cartografiar la actividad de cada neurona del cerebro humano utilizando Big Data, con la esperanza de conocer la "organización del cerebro y comprender los mecanismos de la cognición, el aprendizaje o la plasticidad"[6]. Lo que vemos aquí es un ejemplo claro de esta mímesis que se da entre la IA y las (neuro)ciencias cognitivas: por un lado, el cerebro es imaginado como un dispositivo computacional, pero al mismo tiempo el objetivo es crear nuevo *hardware* con base en el modelo del cerebro, representado por esta tecnología.

Fruto de todo ello es una renovada arquitectura computacional que suele recibir la denominación de chips neurosinápticos o neuromórficos, que descansan en el lenguaje de la neurociencia para desarrollar sistemas que emulan la conectividad del sistema neuronal, trascendiendo la arquitectura clásica de von Neumann. La fusión entre neurociencia computacional e ingeniería cibernética da paso así a lo que los informáticos de IBM han llamado "Cognitive Computing", una Computación cognitiva cuya finalidad es "desarrollar un mecanismo coherente, unificado y universal inspirado en las capacidades de la mente"; lo que buscan es "implementar una teoría computacional unificada de la mente" (Dharmendra, 2011, 62). La imagen del cerebro como una máquina y de la mente como un conjunto de procesos concebidos en términos de automaticidad se refleja aquí de forma transparente, pues lo que aquí se busca es "descubrir, demostrar y entregar los algoritmos centrales del cerebro y obtener una profunda comprensión científica de cómo la mente percibe, piensa y actúa" (Dharmendra, 2011, 65). Es decir, si finalmente descubrimos cómo opera la mente, qué sea la inteligencia, es porque encontramos los algoritmos que la conducen. La visión mecánica del pensar humano es aquí evidente, reducido el pensar humano al esquema algorítimico.

De sumo interés en relación con el chip neuromórfico es que lo que simulan es lo que en la neurociencia se denomina la plasticidad cerebral, tema a

[5] "En contraste con el ámbito experimental, la neurociencia computacional trata de especular "cómo" funciona el cerebro (...) Los estudios dentro de la neurociencia computacional también pueden ayudar a desarrollar aplicaciones como el análisis avanzado de datos de imágenes cerebrales, aplicaciones técnicas que utilizan cálculos similares a los del cerebro y, en última instancia, un mejor tratamiento de los pacientes con daño cerebral y otros trastornos relacionados con el cerebro": Alac, 2011, 3.

[6] https://www.humanbrainproject.eu/en/about/overview/

su vez muy en boga en la neurología de las últimas décadas (aunque su conceptualización moderna hay que situarla en el giro al siglo XX: Bates, 2021). La plasticidad del sistema nervioso central, la plasticidad nerviosa o neuronal, la plasticidad sináptica…de diversa forma se hace referencia a esta facultad de adaptación y evolución que muestra el cerebro, a su estructura siempre en evolución en la fase adulta y su capacidad de reorganizarse tras un traumatismo importante[7]. Y más allá de todas las implicaciones técnico-biológicas que despierta la plasticidad cerebral (sobre las que volveremos *infra*: V), lo que me interesaría aquí subrayar de esta analogía es por un lado la radical humanización de la máquina, que puede ahora en virtud de la absorción de la plasticidad presentar una naturaleza "creativa" e impredecible -se nos dice-, escapando al automatismo asociado a su propia naturaleza; y por otro lado entender que la recepción de la plasticidad cerebral en el discurso filosófico-tecnológico abre vías de reflexión de importancia enorme en esta nueva era digital, pues bien mirado el tema de la plasticidad inherente al cerebro nos plantea, una vez más, pero con un acento mayor y diferente, la cuestión sobre el modo en que las tecnologías digitales pueden no solo influir en nuestro cerebro, sino además -tal vez- modificarlo.

3. INTELIGENCIA NATURAL - INTELIGENCIA ARTIFICIAL

Que la estructura de nuestra inteligencia está cambiando con las tecnologías digitales es algo que viene preocupando a pensadores de diverso signo de un tiempo a esta parte. La proposición aquí a examen es la de que pensamos con y a través de las tecnologías digitales. La tesis la encontramos ya debatida en los textos de Marshall McLuhan (1964), Friedrich Kittler (1992) o Lev Manovich (2002)[8]. La Inteligencia artificial, en este sentido, viene a significar que nuestra inteligencia está *artificialmente* generada: no se trata de la visión tradicional que busca emular la inteligencia humana a través de la máquina artificial, sino de entender la mente humana artificialmente estimulada; no se trata por tanto de la humanización de la máquina, sino de la tecnologización de la mente, y esto a través del constante y regular contacto con el dispositivo digital, que atraviesa nuestro comportamiento diario - conocemos el paulatino sometimiento de nuestras vidas al algoritmo con el cambio de estatus de la tecnología digital, que avanza en la gestión de datos y determinación de conductas a una velocidad más alta de lo que nuestras habilidades humanas son capaces de reconocer[9].

[7] "La plasticidad en el sistema nervioso es una alteración de la estructura o la función provocada por el desarrollo, la experiencia o las lesiones" (Gregory, 1987, 623)

[8] Puede añadirse a la larga lista Clark, 2008; Carr, 2010; Brockmann, 2011; Hayles, 2012.

[9] Para muchos esta relación cotidiana con el algoritmo predictivo representa una especie de antihumanismo radicial. Vid. en este sentido el trabajo reciente de Éric Sadin (2020). Nicholas

(…)

Tomarse en serio la IA en este sentido de la tecnologización de la mente requeriría volver a examinar una vez más la realidad de la cognición humana. Examinar cómo usamos nuestras capacidades cognitivas, pero también cómo las adquirimos en cada caso. Y parece claro que cuanto más se trabaja con tecnologías digitales más se aprecia la capacidad de las máquinas para llevar a cabo tareas cognitivas avanzadas. En este sentido los instrumentos tecnológicos dejan de ser meros instrumentos para convertirse en extensión de los propios pensamientos; el objeto útil deja de ser solo soporte (el teclado, por ejemplo) y pasa ser una forma ampliada de cognición, en una visión dialéctica entre agencia y pensamiento que trasciende la tradicional concepción autónoma de la mente para verla ampliada gracias a la incorporación del artefacto digital en sus procesos de cognición. El esquema seguido por el modelo de la "tecnogénesis" respondería en parte a esta manera "híbrida" de ver las cosas (Hayles, 2012), que en términos históricos no es desde luego nueva. La idea de que los humanos coevolucionaron con el desarrollo de las herramientas no es un asunto debatido entre paleoantropólogos. Ya Ortega nos habló en su *Meditación de la técnica* (1931) de la condición de extraño centauro ontológico que tiene el ser humano, a un tiempo natural y extranatural. "Las manos y los pies -decía John Dewey-, los aparatos y los dispositivos de todo tipo forma parte de él (el pensamiento) tanto como los cambios en el cerebro" (Dewey, 2002, 8). Bajo esta perspectiva la mente es vista en contacto profundo con "las cosas" (un interés por "volver a las cosas" como el que buscaba Husserl en su proyecto de reducción fenomenológica, suspendiendo el juicio con el *epojé* existencial, respondería en parte a esta visión amplia de la mente en su interacción con la realidad)[10]. La percepción, el aprendizaje, el pensamiento y en general el mundo de los sentimientos estarían informados por nuestras interacciones corporales con el mundo que nos rodea. Modernamente se habla de la perspectiva de la corporeidad o encarnación material del pensamiento (*embodied perspective*)[11]. Para Esther Thelen, conocida defensora de esta perspectiva, "la cognición depende de los tipos de experiencias

Carr argumenta en su *The Shallows: What the Internet Is Doing to Our Brains* que los cambios de esta era digital vienen a disminuir nuestra capacidad de concentración, lo que a su vez propicia el pensamiento superficial y en general la disminución general de la capacidad intelectual. Una mente "vagabunda" (*mind wandering*) caracteriza la mente contemporánea, incapaz de consagrar su atención de forma duradera en la realización de una sola actividad, llevada por el impulso de consultar las muchas aplicaciones que reclaman su atención en los dispositivos electrónicos que le acompañan (Carr, 2010). Se insiste así en la disciplina de la atención como la gran tarea pedagógica de nuestro tiempo.

[10] Una apuesta por los métodos de la fenomenología desde la neurociencia es el impulsado por el biólogo y neurólogo chileno Francisco Varela (1999), creador junto con Humberto Maturana del concepto de *autopoiesis*, que después impregnó diversas ramas del saber humanístico, entre otras la del derecho (de que son mejor expresión los trabajos de Niklas Luhmann y Gunther Teubner).

[11] Uno de los primeros trabajos en proponer el enfoque de la cognición *embodied* en las ciencias cognitivas fue el trabajo colectivo del mencionado neurólogo chileno Francisco Varela (1991).

que se derivan de tener un cuerpo con capacidades perceptivas y motoras particulares que están inseparablemente vinculadas y que juntas forman la matriz dentro de la cual se engranan la memoria, la emoción, el lenguaje y todos los demás aspectos de la vida" (2000, 4).

La perspectiva *embodied* recuerda en parte las tesis de Lévi-Strauss *La Pensée Sauvage* (1962), *La domesticación del pensamiento salvaje* (1977) de Jack Godoy o la idea de Hutchins en su *Cognition in the Wild*, para el que "la cognición humana no sólo está influenciada por la cultura y la sociedad, sino que es en un sentido muy fundamental un proceso cultural y social" (Hutchins, 1995: 14). Aquí la cognición humana se nos revela como algo que trasciende siempre al individuo, entendido el proceso mismo de cognición como un fenómeno cultural. Y en esta misma línea de pensamiento, que podríamos decir representa un cierto agnosticismo respecto de la naturaleza de la cognición humana, de nuestra capacidad para teorizar sobre la mente y las facultadas "naturales", encontramos los trabajos de Bruno Latour, que cobran especial relevancia en la reflexión sobre la proyección del fenómeno tecnológico en la vida inteligente "natural". Para Latour los procesos mentales son de facto generados como resultado de la participación del individuo en una red de relaciones, en un proceso que trasciende las facultades cognitivas y que además conoce instancias humanas y no humanas. Por ello el pensador francés se ha mostrado muy crítico con la noción de inteligencia artificial. Con base en su propia "teoría del actor-red" (2005) ya venía argumentando que no hay una inteligencia natural como tal: la distinción entre una inteligencia natural de otra artificial se difumina, pues "con la introducción de tantas tecnologías intelectuales, desde la escritura hasta los laboratorios, desde las reglas hasta los guijarros, desde las calculadoras de bolsillo hasta los entornos materiales, se ha desdibujado la propia distinción entre las inteligencias naturales, situadas y tácitas, y las artificiales, transferibles y desencarnadas" (Latour, 1995, 300-1).

Lo que se aprecia en estas aproximaciones al problema de la tecnología en la era digital es la paulatina difuminación del concepto tradicional de conciencia propia del sujeto individual. Diluida en una red de variables tecnológicas, sociales y neurológicas y respondiendo en su configuración al esquema de la computación -siguiendo así el modelo de la máquina-, la mente humana ha perdido en su automaticidad cualquier significado de autonomía. Además, la plasticidad del cerebro, ahora absorbida por la máquina con miras a la integración de lo contingente, en realidad no altera las cosas. Lejos de humanizar la máquina, volviéndola por fin "inteligente"[12], lo que se hace aquí

[12] Catherine Malabou en cambio quiere pensar que si no se vuelve inteligente, al menos la máquina y el organismo encuentran en el punto concreto de la plasticidad el engarce dialéctico que las ata, disolviendo al fin sus diferencias; esta postura de la pensadora de la plasticidad ha

(...)

es reproducir el patrón de la visión "mecánica" de la mente, proyectando el automatismo de la máquina -herencia del camino paralelo seguido por las ciencias cognitivas y la tecnología- en el proceso mismo de reconstrucción plástica del cerebro, que parece seguir así unos códigos pre-determinados. Pues el nuevo diseño del chip neuromórfico en realidad concibe esta plasticidad como una tecnología automática más: son chips que aprenden, y aunque en este sentido el proceso es más complejo porque integra una causalidad circular que permite a la máquina en cierto sentido su autodeterminación, sigue siendo un aprendizaje desarrollado con base en precisas operaciones algorítmicas preestablecidas. Es decir, no escapa del automatismo. Como George Canguilheim argumentó en su conferencia de 1947 *Máquina y organismo*, las máquinas tienen unos fines a que atienden y que gobiernan su diseño, pero a diferencia de los organismos, la orientación teleológica de la máquina está siempre dada desde fuera. Su propósito no le es inherente (Canguilheim, 1998, 90; interesa al respecto Dreyfus, 1993).

4. PROYECCIÓN DE AUTOMATISMO: EL INCONSCIENTE COGNITIVO

La discusión encendida que se mantiene en la filosofía de la tecnología sobre la tecnologización de la mente y humanización de la máquina, que en muchos casos proyecta el automatismo de la tecnología digital a la conducta humana como consecuencia de la asimilación del lenguaje de la inteligencia artificial, fundida ya en natural en la mirada de Latour, tiene también un interesante punto de inflexión en la discusión abierta en psicología cognitiva sobre la automaticidad, el automatismo y la nueva concepción de la mente humana. Una mente que para una corriente dominante de la psicología y neurología es vista esencialmente como un sistema inconsciente que conduce y determina nuestras acciones sin ninguna intervención de la conciencia: como un autómata. El llamado "nuevo inconsciente" (Uleman, 2005) de la contemporánea ciencia cognitiva subraya esto mismo: la automaticidad inconsciente que gobierna mucho, si no todo, de nuestro aparato mental en sus muy diversos y complejos procesos. Hasta los pensamientos más singulares y aparentemente irreductibles a procesos habituales, como el aspecto creativo, la intuición, el hallazgo, el juicio moral… son tenidos como emanación del inconsciente, accesible ahora gracias al examen que nos ofrece la psicología experimental y el escaneo del cerebro. Como decíamos arriba, la mente, la inteligencia, adopta las características de la tecnología que lo representa, en este caso la electroencefalografía (EEG) y la resonancia magnética funcional (fMRI). Unas herramientas digitales que se ponen en práctica para el estudio de todo

cambiado en el último tiempo con ocasión de la aparición de las nuevas máquinas digitales contemporáneas, concretamente del neurochip (2019, 83). Interesa cf. con la posición contraria que mantenía en 2008.

tipo de actividad mental, como el llamado momento "Aha", o la intuición feliz del genio que tras trabajo y paciencia se ve iluminado con el hallazgo de la idea, la fórmula. Arquímedes, Newton, Beethoven no serían tan genios en esta visión de las cosas, que parece no ver en ellas más que una actividad cognitiva del inconsciente ahora mapeada e interpretada como una especie de automatismo. "Aunque la experiencia de *insight* (o hallazgo) es repentina y puede parecer desconectada del pensamiento inmediatamente anterior, estos estudios demuestran que el *insight* es la culminación de una serie de estados y procesos cerebrales que operan en diferentes escalas de tiempo", de suerte que "los *insights* se producen cuando se computa una solución de forma inconsciente y posteriormente emerge a la conciencia de forma repentina" (Kounios, Beerman 2009, 210). Los procesos inconscientes desempeñan así un papel decisivo en la consecución de las ideas creativas. Los estudios de neuroimagen del cerebro durante el "REST" (Random Episodic Silent Thought, o pensamiento silencioso episódico aleatorio, también denominado estado por defecto) sugieren que las cortezas de asociación son las principales áreas que están activas durante este estado y que el cerebro se reorganiza espontáneamente y actúa como un sistema autoorganizado. Lo que aprecian estos investigadores es que "el proceso creativo se caracteriza por destellos de visión que surgen de las reservas inconscientes de la mente y el cerebro" (Andreasen, 2011).

Este nuevo inconsciente hay que diferenciarlo del inconsciente de Freud, que ha definido el concepto en el último siglo[13]; lejos de deseos del irracional, típicamente primarios, este "nuevo inconsciente" es coherente, sistemático y racional. Es el "inconsciente cognitivo", como lo llamó Kihlstrom, que fue quien primero lo describió (1987). Y lo que más nos interesa es que allí Kihlstrom ya describía las formas en que la computadora como metáfora servía la base para una concepción cada vez más compleja de los procesos mentales. Hoy este inconsciente sigue siendo básicamente cognitivo, e históricamente ligado al ordenador como metáfora: "La metáfora del ordenador legitimó teorías complejas sobre procesos inobservables en su conjunto, evitando aparentemente los pecados de antropomorfizar", dicen los autores del *New Unconscious* (Uleman, 2005, 4). El funcionamiento del inconsciente es rastreado a través de métodos y tecnologías que siguen la trazabilidad de la lógica computacional. Una vez más, es el modelo de la inteligencia artificial el aquí seguido, un modelo que, así como es capaz de imitar -en diverso grado- al humano en diversas actividades (y en su imitación operar de forma automática -y por tanto inconsciente-), del mismo modo se piensa que el resto de actividades mentales más complejas, de poderse igualmente emular, entonces su naturaleza será igualmente inconsciente, pues una máquina es capaz de

[13] Una historia autorizada del inconsciente es la que presentó Henri Ellenberger (1970).

reproducirlas. Lo que vemos aquí es una ilusión de consciencia, diluida por obra del enfoque automatista de la máquina, que todo lo impregna. La clave del análisis, una vez más, es la tecnologización de la mente, pues no es la máquina la que cobra *conciencia* en esa búsqueda sin fin de la inteligencia artificial "general", sino que es la mente humana la que es cifrada en términos mecánicos gracias a la reproducción del modelo cerebral. Un modelo de enorme complejidad, pero reducido aquí a conexiones que escapan a lo que denominamos conciencia.

5. LA "EDAD DEL AUTÓMATA" O LOS ORÍGENES DEL AUTOMATISMO

Más allá del beneficio que pueda o no tener este tipo de enfoque "computacional" en la investigación científica que representa el amplio arco de las ciencias cognitivas, lo que interesa subrayar es la distancia cada vez mayor que separa al ser humano de aquello que mejor le caracteriza: su mente inteligente. Si nos preguntamos cómo hemos llegado a este punto del pensamiento autómata y el inconsciente, la historia que habríamos de trazar se remontaría al umbral de la computación moderna y de la Inteligencia artificial, que es parejo al de las ciencias cognitivas (Edwards, 1996, 239 y ss.), como decíamos al principio de este texto. Pero esta historia a su vez participaría de otra mayor, aquella que se pregunta por las formas en que la mente se ha visto ligada en su concepción a las diversas tecnologías desde la revolución científica del siglo XVII. En la senda abierta por Descartes, cuyos escritos filosóficos y fisiológicos (es decisivo aquí su *Tratado sobre el hombre*, 1662) nos dieron la visión del moderno autómata -es decir de una máquina pensante-, encontramos a Spinoza haciendo esfuerzos por sistematizar la mente racional, un intelecto puro que con base en el moderno dualismo sustancial escapaba a toda explicación corporal y así de cualquier expresión de automaticidad. Pero el alma, en tanto "autómata espiritual" (Marshall, 2014), no quedaba fuera de las leyes mecánicas del universo, y expresaba así un conjunto coherente de operaciones (un mundo que se convertía paulatinamente en objeto de la naciente psicología). Para Leibniz, "la operación de los autómatas espirituales, es decir, de las almas, no es mecánica; pero contiene eminentemente lo que hay de precioso en la mecánica; los movimientos desenvueltos en los cuerpos estan reconcentrados allí por la representación, como en un mundo ideal, que explica las leyes del mundo actual y sus consecuencias" (2021, § 406). Ese mundo intelectual puro dibujado por el racionalismo cartesiano que dividió el fenómeno intelectual del de la sensibilidad corporal lo trascenderían pensadores de la ilustración para examinar cómo las ideas y la experiencia formaban gradualmente un sistema autónomo, y lo que vemos en Hume es la naturalización de la razón: para el inglés, "la razón no es más que un instinto

maravilloso e inteligible en nuestras almas, que nos lleva por un determinado tren de ideas" (Hume, 1888, 179); es decir, una tendencia interna.

Por otra parte, las teorías de la cognición "autómata" están atadas al desarrollo de máquinas inteligentes en este tiempo de luces. La posibilidad misma de una inteligencia artificial había sido ya dibujada por pensadores como Pascal y Leibniz en la invención de la calculadora semi-automática. Por esa ambición típica del pensamiento ilustrado de descifrar el mecanismo inherente de toda realidad se habla también de "la edad del autómata" (como hace Schaffer, 1999) para referirse a este tiempo de luces, un tiempo que imaginó el autómata como una máquina con forma humana y como un humano que obra como la máquina. Ciertamente, en la larga historia de intentos de crear autómatas artificiales, que se remonta a la antigüedad y llega hasta nuestros días, el siglo XVIII ocupa un lugar destacado en la fabricación de androides[14]. Los tres autómatas fabricados por Jacques de Vaucanson en la década de 1730 (flautista, tamborilero y galoupet) figuran entre los casos históricos más conocidos, suscitando gran interés entre filósofos y científicos en la época. Captaron la atención de Voltaire, que celebró a su inventor como "el rival de Prometheus" y persuadió a Federico el Grande para que invitara a su fabricante a unirse a su corte (Voltaire, 1738, 420) -finalmente fue Luis XV quien lo hizo (Riskin, 2003, 601). Del siglo XIX es la primera tecnología de pensamiento completamente automático, diseñada por Charles Babbage -la Máquina Analítica de la década de los 40. De los 60 es el "piano lógico" de William Jevons, un instrumento que desplegaba los nuevos desarrollos del álgebra de Boole para la resolución de cuestiones lógicas. No sorprende entonces que se estableciera en este contexto una analogía contemporánea entre de una parte los procesos automatizados y mecanizados de estos autómatas y el funcionamiento de la cognición humana entendida, así como una entidad de naturaleza inconsciente y automática. Thomas Henry Huxley, el principal darwinista británico, decidió que la mente, al igual que otros aspectos de la vida animal, era la consecuencia directa de la maquinaria animal, y que los animales y los seres humanos eran esencialmente "conscientes, sensibles, autómatas" (Huxley, 1899, 238).

En respuesta a la tesis del ser humano como autómata, defendida entre otros por William James Huxley, eminente psicólogo de finales del XIX (además

[14] Importa aquí referir los trabajos de Alfred Chapuis, quien presentara en los años veinte del siglo pasado uno de los primeros trabajos sobre la historia de los autómatas mecánicos que aún hoy sigue siendo de los más completos, y que ha informado los muchos estudios que vienen realizándose sobre el particular. Vid. Chapuis/Gerlis, 1928 (los capítulos 1- 4 para el autómata antiguo); Chauis/Droz, 1958. Puede añadirse Beyer, 1983; Bailly, 1987; Beaune, 1990, y el reciente volumen colectivo Cave *et al.*, 2020. Para la antigüedad cf. el trabajo reciente de Mayor, 2020.

de filósofo fundador del pragmatismo y hermano del escritor Henry James) escribió el famoso ensayo "Are we Automata?" (1879), que resultaría decisivo para abrir las puertas a nuevas formas de entender la psicología más allá de los corsés estrechos en que la había situado el cientificismo del XIX[15]. Basándose en las teorías neurológicas de su tiempo, James trascendió la visión de la mente asociada al automatismo. La vio como un especial tipo de autómata, de un lado estable pero libre y abierta a lo nuevo de otro. Interesado en la plasticidad cerebral, situó tanto la automaticidad como su interrupción en el cerebro (Bates 2021: 117-122). De este modo ayudó a considerar la extensión como parte del pensamiento, superando la mirada moderna que la veía como algo inerte y al margen del pensamiento. Como amigo de Bergson, le fascinó la idea de una energía creadora capaz de expresar un impulso vital (*élan vital*) que se sitúa más allá de todo determinismo, y que se desarrolla continuamente generando nuevas formas creadoras. Borges, al memorar su figura en una Nota preliminar a su *Pragmatismo*, que es pieza universal de la historia de la filosofía, decía de James que en la lucha entre aristotélicos y platónicos, estos que intuyen ideas y aquellos que solo ven generalizaciones, aparecía la figura de James para enriquecer "esa lúcida tradición. Como Bergson, lucha contra el positivismo y contra el monismo idealista. Aboga, como él, por la inmortalidad y la libertad" (Borges, 1945, 10)[16].

Esta dialéctica que se expresa en la resistencia del automatismo a sí mismo es evidente desde los primeros momentos de la historia de la inteligencia artificial. Dos direcciones en conflicto dirigían la primera IA: de un lado, los científicos de la computación que consideraban el cerebro como una máquina de aprendizaje algorítmico más o menos complejo pero cifrable en sus procesos automáticos y por tanto reproducible por ordenadores; de otro lado un grupo no menor de investigadores pioneros de la cibernética reconocieron la importancia que tenía la plasticidad cerebral para el desarrollo de la IA, pues esta les permitía modelar la parte creativa e impredecible de la inteligencia humana, un proceso abierto que iba más allá del automatismo típico de los procesos mecánicos habituales. Turing, Ashby y Von Neumann, personajes

[15] Los progresos crecientes de la Física y más tarde de la Biología concebían al ser humano sin más como un conjunto de sensaciones, voliciones, juicios, reducibles a sensación, impulso de vivir, apetito, en la jerga de la época; el libro de Bergson *La evolución creadora* (1907) se presentó como el ensayo filosófico de la cuestión de la inteligencia que pretendía liberarla de la prisión teórica en que la tenía fijada la psicología positiva.

[16] Prosigue Borges: "El universo de los materialistas sugiere una infinita fábrica insomne; el de los hegelianos, un laberinto circular de vanos espejos, cárcel de una persona que cree ser muchas, o de muchas que creen ser una; el de James, un río. El incesante e irrecuperable río de Heráclito. El pragmatismo no quiere coartar o atenuar la riqueza del mundo; quiere ir creciendo como el mundo" (1945: 12).

fundamentales de la historia de la cibernética, se interesaron por el cerebro plástico en la creencia de que les ofrecería la posibilidad de simular los saltos creativos e imprevisibles de la inteligencia humana, unos procesos que estaban más allá de la ejecución de un programa automático. Las similitudes con el tipo de trabajo que venían desarrollando al respecto eminentes psicólogos, psiquiatras y neurólogos de aquellas décadas es elocuente: no solo William James, también Karl Lashley, Kurt Golstein o Wolfgang Köhler, entre otros, subrayaban la capacidad del sistema nervioso para repararse y reestructurarse a sí mismo. El cerebro plástico era imaginado como un sistema abierto que se autodeterminaba pero que también era capaz de desafiar su propia automaticidad. Y los creadores del ordenador digital se inspiraron explícitamente en este concepto neurofisiológico de plasticidad en su ambición de simular la mente inteligente (Bates, 2016). Como consecuencia del examen de la plasticidad, que se expresa en saltos imprevisibles de la inteligencia capaces de ir más allá de la ejecución automática de comportamientos habituales, la neurofisiología de la época mostró un interés natural en los trastornos y crisis del cerebro lesionado. Una patología cerebral que atrajo también a la cibernética de la época, convencida de que estos momentos disruptivos esenciaban la inteligencia humana a simular. Así, en su clásico de *Cybernetics: Or, Communication and Control in the Animal and Machine* (1948), Norbert Wiener afirmaba que ciertas inestabilidades psicológicas tenían analogías tecnológicas bastante precisas: "Los procesos patológicos de naturaleza algo similar no son desconocidos en el caso de las máquinas de computación mecánicas o eléctricas" (Wiener, 1948, 172). El interés de la cibernética por las formas patológicas de cuerpo y mente no sorprende, considerando que muchos de ellos habían sido también profesionales de la medicina. Warren McCulloch fue neurólogo y trabajó en clínica psiquiátrica, Ashby era psiquiatra y Rosenblueth cardiólogo. En cierto modo la cibernética se convirtió en una disciplina medicalizada, con el foco puesto en identificar el origen de la inestabilidad en sistemas complejos para eliminarlas y recuperar la estabilidad (Pickering, 2010, 91 y ss.). "La cibernética -son palabras de Ashby- ofrece la esperanza de proporcionar métodos eficaces para el estudio, y el control, de sistemas que son intrínsecamente muy complejos.... De este modo, ofrece la esperanza de proporcionar los métodos esenciales para atacar los males -psicológicos, sociales, económicos- que actualmente nos derrotan por su complejidad intrínseca" (Ashby, 1961, 5-6; Cf. Bates, 2014, 33).

Este breve esbozo histórico nos permite apreciar la complejidad que arrastra la idea del autómata, con frecuencia obviada por los modelos tecnológicos contemporáneos, en exceso simplistas cuando equiparan la inteligencia humana y su inconsciente con lo automático. En la relación recíproca y de espejo que guardan el cerebro y la computadora, el aparato

mental no puede verse como mero resultado de un conjunto automatizado de procesos cognitivos, sino como expresión de una indeterminación previa, fruto de una plasticidad cerebral que le es inherente y que va más allá de la ejecución automática de comportamientos. Las consecuencias filosóficas para la idea de autonomía, que no se ve así vaciada de significado como consecuencia de la automaticidad, son decisivas. En palabras de Malabou, filósofa interesada en la plasticidad: "La plasticidad destructiva despliega su obra a partir del agotamiento de las posibilidades, cuando toda virtualidad se ha ido desde hace tiempo […] al instaurar la relación entre el ser y el accidente fuera de todo concepto de la predestinación psíquica, al marcar la importancia del surgimiento brutal e inesperado de la catástrofe, no me convierto en la guardiana de un pensamiento puro ni de una idolatría de la sorpresa. Por el contrario, me rehúso a considerar que el accidente responde al llamado de una identidad que, en cierto sentido, solo lo esperaría para desplegarse. Definitiva y resueltamente, sé que «es peligroso buscar la esencia»" (Malabou, 2008, 72-73). Pero esa esencia de la inteligencia se viene persiguiendo con el mismo tesón positivista, aun en estos tiempos líquidos y posmodernos, y la calculabilidad añadida por el matematizante mundo de la razón a esa esencialidad en examen ofreció y aún ofrece un resultado pobre del concepto de inteligencia, ciertamente abstracto y vacío (del que con razón se quejaba María Zambrano en *Hacia un saber sobre el alma*; por ejemplo, en 1987, 105).

6. NOTA FINAL

Desde este legado de la cibernética y la IA que hemos tratado de dibujar en paralelo con las ciencias cognitivas, podemos replantear críticamente la cuestión arriba mencionada referida al modo como las tecnologías digitales están influyendo nuestro cerebro, en una concepción de la inteligencia artificial que pierde su significado originario de la "máquina inteligente" para incidir en cambio en el modo como la mente viene artificialmente estimulada desde los múltiples aparatos digitales que como prótesis la acompañan. El pensador de la tecnología Bernard Stiegler advertía la dialéctica peligrosa del automatismo, que es la esencia de la tecnología digital: "La automatización hace posible la digitalización, pero aunque aumenta inconmensurablemente el poder de la mente (como racionalización), también puede destruir el conocimiento de la mente (como racionalidad). Un pensamiento «farmacológico» de lo digital debe estudiar las dimensiones contraproducentes de la automatización para contrarrestar sus efectos destructivos sobre el conocimiento" (Stiegler, 2018, 38). En efecto: lo que vemos es que la tecnología por un lado aporta un enorme conocimiento y desarrollo de la inteligencia a través de lo digital, que implica una suerte de exteriorización enorme de la memoria -entre otras muchos factores que aceleran y modifican nuestra percepción cognitiva-, pero por otro lado puede arruinar la capacidad crítica del ser humano y caer en la

automaticidad de su propia autodeterminación -entre otras razones por verse conducido de forma exponencial por la predicción algorítmica, que condiciona y a menudo determina cada vez más de sus decisiones tanto de su parcela privada como de la vida pública (Huergo Lora, 2020). Se llega así peligrosamente a la automaticidad vaciada de toda autonomía. Y ciertamente "sería un día triste si los seres humanos, al adaptarse a la revolución informática, se volvieran tan perezosos intelectualmente que perdieran el poder del pensamiento creativo" (Gardner, 1978, VI-VIII).

La digitalización de prácticamente todas las esferas sociales e individuales permite hablemos del automatismo como el signo o *Geist* de nuestro tiempo, en la conciencia cada vez más clara de que habitamos una "sociedad automática", por hablar de nuevo con Stiegler (2016). En este contexto resulta de enorme interés y actualidad las críticas de la razón moderna que encontramos, por ejemplo, en los teóricos de la Escuela de Frankfurt, que tanto han interesado al pensamiento crítico de posguerra. Estos pensadores estaban preocupados -precisamente en este tiempo de nacimiento de la cibernética- en el peligro de que la inteligencia se instrumentalizara cayendo sin remedio en el automatismo. En *El eclipse de la razón* Horkheimer adoptaba un tono grave que hoy podemos acoger nosotros para cerrar estas reflexiones: "cuanto más se han automatizado las ideas, se han instrumentalizado, menos se ven en ellas pensamientos con significado propio. Se consideran cosas, máquinas. El lenguaje se ha reducido a una herramienta más del gigantesco aparato de producción de la sociedad moderna" (Horkheimer, 2004, 15).

7. BIBLIOGRAFÍA

Alac, Morana (2011), *Handling Digital Brains. A Laboratory Study of Multimodal Semiotic Interaction in the Age of Computers*, MIT Press, Cambridge.

Andreasen, Nancy C. (2011), "A Journey into the Chaos: Creativity and the Unconscious", en: *Mens Sana Monographs* 9, 42-53.

Ashby, W. R. (1948), "Design for a Brain", en: *Electronic Engineering*, 20 (December), 379-383.

Bailly, Christian (1987), *Automata. The Golden Age*, P. Wilson, for Sotheby's, London.

Bates, David (2014), "Unity, plasticity, catastrophe: Order and pathology in the cybernetic era", en: Lebovic, Nitzan, Andreas Killen (eds.), *Catastrophes A History and Theory of an Operative Concept,* Oldenburg, De Gruyter, 32-54.

— (2016), "Automaticity, Plasticity, and the Deviant Origins of Artificial Intelligence," en: *Plasticity and Pathology: On the Formation of the Neural Subject*, New York, Fordham University Press, 194-218.

— (2021), "Unstable Brains and Ordered Societies: On the Conceptual Origins of Plasticity, ca. 1900", en: Natasha Lushetich, Iain Campbell eds., *Distributed Perception Resonances and Axiologies*, London, Routledge, 117-129.

Beaune, Jean-Claude (1989), "The Classical Age of Automata: An Impressionistic Survey from the Sixteenth to the Nineteenth Century," en: Feher, Michel (ed.), *Fragments for a History of the Human Body*, I, Cambridge, MIT Press/Zone, 430-480.

Beyer, Annette (1983), *Faszinierende Welt der Automaten: Uhren, Puppen, Spielereien*, Callwey, München.

Borges, Jorge Luis (1945), "Nota preliminar", en: James, W., *Pragmatismo. Un nombre nuevo para algunos viejos modos de pensar*, Emecé, Buenos Aires.

Brockman, John (ed.) (2011), *Is the Internet Changing the Way You Think? The Net's Impact on Our Minds and Future*, Harper, New York.

Canguilhem, George (2008), "Machine and Organism," en: *Knowledge of Life*, Geroulanos, Stefanos, Daniela Ginsburg (trad.), New York, Fordham University Press, 75-97.

Carr, Nicholas (2010), *The Shallows: What the Internet Is Doing to Our Brains*, W. W. Norton, New York.

Cave, Stephen *et al.* (eds.) (2020), *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*, Oxford University Press, Oxford.

Chapuis, Alfred, Edouard Gelis (1928), *Le monde des automates: Étude historique et technique*, 2 vols., Chez les auteurs, Paris.

Chapuis, Alfred, Edmond Droz (1958), *Automata: A Historical and Technological Study*, Éditions du Griffon, Neuchâtel.

Clark, Andy (2008), *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*, Oxford University Press, London.

Dewey, John (2002), *Essays in Experimental Logic*, ed. Micah Hester, Robert B. Talisse, Southern Illinois University Press, Carbondale.

Dharmendra S. Modha *et al.* (2011), "Cognitive Computing", en: *Communications of the ACM* 54, 8, 62-71.

Dreyfus, Hubert (1993), *What Computers Still Can't Do. A Critique of Artificial Reason*, MIT Press, Cambridge, MA.

Dupuy, Jean-Pierre (2009), *On the Origins of Cognitive Science: The Mechanization of Mind*, trans. M. B. DeBevoise, MIT Press, Cambridge.

Dumit, Joseph (2004), *Picturing personhood: Brain scans and biomedical identity*, Princeton University Press, Princeton.

Edwards, Paul (1996), *The closed world: computers and the politics of discourse in Cold War America*, MIT Press, London.

Ellenberger, Henri F. (1970), *The Discovery of the Unconscious. The History and Evolution of Dynamic Psychiatry*, Basic Books, New York.

Gardner, Howard (1985), *The Mind's New Science: A History of the Cognitive Revolution*., Basic Books, New York.

Gardner, Martin (1989), *Aha! Insight*, Scientific American, New York.

Goertzel, B., C. Pennachin (2009), *Artificial General Intelligence*, Springer, Berlin-Heidelberg.

Gregory, Richard L. (ed.) (1987), "Plasticity in the Nervous System", en: Gregory, Richard (ed.) *The Oxford Companion to the Mind*, Oxford University Press, Oxford, 83-100.

Hayles, Katherine (2012), *How We Think. Digital Media and Contemporary Technogenesis*, University of Chicago Press, London.

Hanger, Michael (1997), *Homo cerebralis: Der Wandel vom Seelenorgan zum Gehirn*, Berlin Verlag, Berlin.

Heims, Steve Joshua (1991), *The Cybernetics Group*, Cambridge MIT Press, Cambridge.

Kihlstrom, J. F. (1987), "The cognitive unconscious", en: Science, 237, 1445-1452.

Horkheimer, Max (2004), *The eclipse of reason* (1974), Oxford University Press, New York.

Hume, David (1888), *Treatise of Human Nature*, Oxford University Press, Oxford.

Hutchins, Edwin (1995), *Cognition in the Wild*, The MIT Press, Cambridge.

Huxley, Thomas H. (1899), "On the Hypothesis That Animals Are Automata" (1874), en: *Methods and Results: Essays*, D. Applewood, New York, 199-250.

James, William (1879), "Are we Automata", en: *Mind*, 4, 1-22.

Kittler, Friedrich A. (1992), *Discourse Networks, 1800/1900,* Metteer, Michael (trad.), Stanford University Press, Stanford.

— (1999), *Gramophone, Film, Typewriter*. Translated Geoffrey Winthrop-Young. Stanford University Press, Stanford.

Kounios, John, Mark Beerman (2009), "The Aha! Moment: The Cognitive Nerusocence of Insight", en: *Current Directions in Psychological Science*, 18, 210-216.

Latour, Bruno (1996), "Social theory and the study of computerized work sites", en: Orlinokowski, W. J, Walsham Geoff, (eds.), *Information Technology and Changes in Organizational Work*, Chapman and Hall, London, 295-307.

— (2005), *Reensamblar lo social. Una introducción a la teoría del actor-red* trad. Gabriel Zadunaisky, Manantial, Buenos Aires.

Leibniz, Gottfried (2021), *Teodicea*, en: *Obras Completas. Nueva edición integral*, Wisehouse Classics.

Manovich, Lev (2002), *The Language of New Media*, MIT Press, Cambridge.

McCarthy John *et al.* (2006), "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence" (1955), en: *AI Magazine*, 27, 4, 12-14.

Marshall, Eugene (2014), *The spiritual Automaton. Spinoza´s Science of the Mind*, Oxford University Press, Oxford.

Mayor, Adrienne (2018), *Gods and Robots. Myths, Machines and Ancient Dreams of Technology*, Princeton University Press, Princeton-Oxford.

McLuhan, Marshall (1964), *Understanding Media: The Extensions of Man*, Mentor, New York.

Malabou, Catherine (2008), *What should we do with our brains*, Fordham University Press, New York.

— (2018), *Ontología del accidente. Ensayo sobre plasticidad destructiva*, trad. Cristóbal Durán, Pólvora editorial, Santiago de Chile.

— (2019), *Morphing Intelligence. From IQ Measurements to Artificial Brains*, Columbia University Press, Nueva York.

Pickering, Andrew (2010), *The Cybernetic Brain: Sketches of Another Future*, University of Chicago Press, Chicago.

Pinker, Steven (1997), *How The Mind Works*, Penguin Books, New York.

Riskin, Jessica (2003), "The Defecating Duck or The Ambiguous Origins of Artificial Life", en: *Critical Inquiry*, 29, 599-633.

Thelen, Esther (2000), "Grounded in the world: Developmental origins of the embodied mind", en: *Infancy* 1, 1, 3-28.

Trappenberg, Thomas P. (2010), *Fundamentals of Computational Neuroscience*, Oxford University Press, Oxford.

Sadin, Éric (2020), *La inteligencia artificial o el desafío del siglo. Anatomía de un antihumanismo radical*, trad. Margarita Martínez, Caja Negra Buenos Aires.

Schaffer, Simon (1999), "Enlightened Automata", en: Clar, William, Golinski, Jan, Schaffer, Scimon (eds.), *The Sciences in Enlightened Europe*, Chicago, University of Chicago Press, Chicago, 126-165.

Stiegler, Bernard (2002), *La técnica y el tiempo. II: La desorientación*, trad. Beatriz Morales Bastos, Hiru.

— (2013), "Die Aufklärung in the Age of Philosophical Engineering", en: M. Hildebrandt *et al.* (eds.), *Digital Enlightenment Yearbook*, IOS Press, 29-39.

— (2016), *Automatic Society, 1: The Future of Work*, Polity, Cambridge.

Uleman, James S. (2005), "Introduction: Becoming Aware of the New Unconscious", en: *The New Unconscious*. https://nyuscholars.nyu.edu/en/publications/introduction-becoming-aware-of-the-new-unconscious

Hasin, Ran R, James S. Uleman, (ed), *The New Unconscious*, Oxford, Oxford University Press, 3-18.

Varela, Francisco *et al.* (1992), *The Embodied Mind: Cognitive Science and Human Experience*, MIT Press, Cambridge.

— (1999), *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science*, Stanford University Press, Stanford.

Vidal, Fernando, (2009), "Brainhood, anthropological figure of modernity", en: *History of Human Sciences*, 22, 1, 5-36.

— (2011), *The Sciences of the Soul. The Early Modern Origins of Psychology*, The University of Chicago Press, London.

Voltaire, (1877), "Discours en vers sur l'homme", 1738, *Oeuvres complètes*, 9, Garnier, Paris.

Wiener, Norbert (1948), *Cybernetics: Or, Communication and Control in the Animal and Machine*, MIT Press, Cambridge.

Zambrano, María (1987), *Hacia un saber sobre el alma*, Alianza, Madrid.

CAPÍTULO VIII

SINGULARIDAD TECNOLÓGICA, METAVERSO E IDENTIDAD PERSONAL: DEL HOMO FABER AL NOVO HOMO LUDENS

FERNANDO H. LLANO ALONSO[1]
Universidad de Sevilla
llano@us.es

1. INTRODUCCIÓN

Sostenía Ortega y Gasset en su ensayo *Meditación de la técnica* (1931), que el ser humano tiene la extraña condición de ser a un tiempo natural y extranatural; es una especie de centauro ontológico que media porción de él está inmersa en la naturaleza, mientras que la otra trasciende de ella. Con esta metáfora mitológica pretendía ilustrar el pensador madrileño su idea de que el hombre no es una cosa sino una pretensión:

> Cuerpo y alma son cosas, y yo no soy una cosa, sino un drama, una lucha por ser lo que tengo que ser (Ortega y Gasset, 2006b, 571).

Para Ortega, la vida no es algo que a los hombres se les de hecho, regalado, sino algo que ellos mismos deben hacer:

> El hombre, quiera o no, tiene que hacerse a sí mismo, autofabricarse (*Ibid*, 573).

Ortega apunta a una idea del hombre que en realidad va incluso más allá de la noción antigua de *homo faber*, según la cual "el hombre es la medida de todas las cosas"; el concepto moderno de *homo faber* reemplaza, como diría Hannah Arendt, a las nociones clásicas de armonía y sencillez colocando en su lugar la labor, el trabajo, la producción y la acción como elementos esenciales de una vida activa en la que el hombre instrumentaliza el mundo, lo construye y lo transforma con la fabricación de objetos artificiales que le resultan útiles para realizar ese cometido. Paradójicamente, advierte la pensadora alemana, este cambio en la mentalidad del hombre constructor moderno supuso el origen de su derrota al privarle de los modelos que le habían servido como referencia antes de la Era Moderna.

> Quizá nada indica con mayor claridad el fundamental fracaso del *homo faber* en afirmarse como la rapidez con que el principio de

[1] Estudio realizado en el marco del Proyecto de I+D del Ministerio de Ciencia e Innovación de España *Biomedicina, Inteligencia Artificial, Robótica y Derecho: los Retos del Jurista en la Era Digital* (PID2019-108155RB-I00).

> utilidad, la quintaesencia de su punto de vista sobre el mundo, desapareció y se reemplazó por el de "la mayor felicidad del mayor número" (Hannah Arendt, 1959, 281).

Existe una clara relación de identidad entre técnica y bienestar. Una vez superada la etapa moderna del *homo faber*, el hombre contemporáneo no tiene particular interés en estar en el mundo, en lo que tiene especial empeño es en estar bien; es más, de todos los animales, el hombre es el único para el que lo superfluo resulta necesario, y precisamente en esto consiste la técnica: en la producción de lo superfluo (Ortega y Gasset, 2006b, 561-562).

La vida cotidiana de la sociedad de nuestro tiempo comparte caracteres comunes con el sentido de lúdico que tanto se ha desarrollado en la cultura contemporánea, como señala Johan Huizinga. En efecto, el *homo ludens*, que toma el relevo del *homo faber*, se sumerge en una esfera temporal de actividad que tiene una vida y tendencia propia, pero que no es en sí la "vida corriente", es decir, "la vida propiamente dicha". Al contrario, a través de su sentido lúdico y de la realidad alternativa del juego, el hombre parece practicar el escapismo de los asuntos y las cosas que conciernen a su realidad cotidiana (Huizinga, 2010, 21).

A diferencia del *homo faber*, un hombre que fabricaba cosas con partes materiales que ensamblaba para formar un todo armónico (Bergson, 1973, 91), el *homo ludens* no necesita vencer las resistencias de la realidad material mediante el trabajo, su vida no será "un drama que le obligue a actuar, sino un juego" (Han, 2021, 22). El *homo ludens* trasciende la realidad natural, en la que el individuo se adapta al medio, e inventa a través de la técnica una *sobrenaturaleza* en la que es el medio quien se amolda a la voluntad del sujeto.

Sin embargo, hasta tal punto ha llegado a depender el *homo ludens* de la técnica en el desarrollo de su vida cotidiana, y tan desmedida es su fe en la tecnología, que ha terminado desdibujando su propia identidad y vaciando su propia existencia. Ortega anticipó con clarividencia este desvanecimiento ontológico del hombre contemporáneo ante la creciente autonomía de las máquinas con las siguientes palabras:

> la técnica, al aparecer por un lado como capacidad, en principio ilimitada, hace que al hombre, puesto a vivir de fe en la técnica y sólo en ella, se le vacía la vida (Ortega y Gasset, 2006, 596).

En este proceso agónico del espíritu humano, ante su progresivo desplazamiento del escenario de la realidad física por la irrupción de la revolución tecnológica, no solo hace que el hombre renuncie a la condición de artesano o fabricante y su función quede reducida a la de mero auxiliar de la máquina, sino también que se produzca una disociación entre el cuerpo y el

espíritu humano en el universo virtual creado por el individuo como recurso evasivo de la naturaleza. Precisamente en el espacio digital en el que se replica artificialmente la naturaleza, encontrará el hombre su refugio lúdico de imágenes y sensaciones virtuales alejado del mundo de las cosas. A propósito de esa propuesta de disociación entre cuerpo y espíritu humana, facilitada por el avance neotecnológico, se preguntaba Ortega si ese presuntuoso espíritu que pretende emanciparse de la realidad de las cosas que se ven y se tocan no sería más que pura "demencia" (Ortega y Gasset, 2006b, 603).

A propósito de la alienación del hombre en el espacio virtual advierte Byung-Chul Han en su ensayo sobre las *No-cosas* (2021) que, a medida que aumenta el control que ejercen los algoritmos en el desarrollo de la vida cotidiana de los seres humanos, éstos van perdiendo también su autonomía, la libertad de obrar y decidir por sí mismos. En esta segunda fase de la mecanización, las máquinas autómatas ya no son simples herramientas, cosas inertes manejadas por el *homo faber*, sino *infómatas* que actúan y piensan por los hombres. En ese mundo virtual dominado por la Inteligencia Artificial *apática* (sin *pathos*), la información extraída de la minería de los datos (*data mining*), y el conocimiento -que en este caso no es sabiduría- basado en cálculos almacenados en el *Big Data,* representan una realidad inmaterial, la experiencia sin presencia de hombres superficialmente felices, pero abducidos por los dispositivos digitales, como las tablets y los smartphones, en la era del *phono sapiens* (Han, 2021, 18-21).

La absorción del hombre por el universo virtual de las tecnologías digitales, el abandono por su parte del mundo real y de la realidad tangible produce en el individuo una forma de profunda crisis de identidad, una especie de aguda desorientación respecto al lugar en el que se encuentra. Precisamente, a propósito de esa pérdida de orientación del hombre contemporáneo, Charles Taylor sostiene que dicha desorientación equivale a no saber quiénes somos ni a qué lugar pertenecemos; en definitiva, supone no tener identidad o haberla perdido (Taylor, 2020, 53-55).

Mientras van perfilándose paulatinamente las líneas que demarcan el horizonte de la singularidad tecnológica, hipótesis en la cual la creación de IA fuerte a cargo de las máquinas superará supuestamente el control y la capacidad de la inteligencia humana, la identidad de los individuos se va difuminando cada vez más ante ese futuro transhumano en el que, tanto su posición como su papel, son del todo inciertos. La crisis de *identidad personal y humana*[2] ante la progresiva autodeterminación de las máquinas desarrolladas

[2] Como advierte Rafael de Asís, aunque la identidad humana e identidad personal han sido muy relevantes en el proceso de construcción de los derechos humanos, no deben confundirse: la *(...)*

con IA, así como la expansión del mundo virtual y la inmersión del individuo en un metauniverso (o *metaverso*) en una experiencia multisensorial y tridimensional que se disfruta mediante el uso aplicado de dispositivos y desarrollos tecnológicos de internet, plantean innumerables interrogantes de índole antropológico, ético, político y sociológico. Ahora bien, dada la temática específica de este trabajo, el presente capítulo se centrará exclusivamente en algunas cuestiones ético-jurídicas que surgen en la experiencia jurídica digital a raíz de nuevas formas contractuales a través de la tecnología *blockchain* que permiten transacciones con criptomonedas como Bitcoin, la compraventa de activos digitales no fungibles (cuyas siglas en inglés son NFT), o la posibilidad de realizar contratos autoejecutables (*smart contracts*) en los que desaparece la intervención humana en cualquier operación, que no requieren participación jurisdiccional, además de suponer un ahorro en gestiones burocráticas y una automatización de las operaciones.

Además de las múltiples ventajas que la tecnología *blockchain* ofrece al mundo de los negocios jurídicos y de la economía digital, conviene también replantearse en términos iusfilosóficos los efectos producidos por la aplicación de las Nuevas Tecnologías (NN.TT.) en el ámbito de los derechos y libertades de los individuos (por ejemplo, en relación con la protección de datos o el derecho al olvido). Esta circunstancia hace necesaria la implementación de un marco jurídico digital que proporcione a los usuarios el disfrute de las herramientas que ponga a su alcance una Inteligencia Artificial cada vez más fuerte, pero que también sea más fiable y segura.

A propósito de las implicaciones ético-jurídicas surgidas a partir de la interacción entre el *novo homo ludens* con el metauniverso de internet y la tecnología de la IA sería oportuno determinar en qué medida se está produciendo no solo la desnaturalización del hombre contemporáneo, sino también, en cierto modo, la deshumanización de la técnica en aras de un salto evolutivo que, como pronostican los transhumanistas, nos acerque como especie al horizonte de singularidad del *homo excelsior* (híbrido entre hombre y máquina inteligente), todo ello sin que sirva de excusa para soslayar los beneficios y el bienestar que la revolución tecnológica 4.0, y en particular la IA y

identidad personal se expresa en forma de "condición personal (percepción, voluntad, imaginación, memoria, intuición, razón y los órganos que se soportan) y situación personal (contexto). Y, además, presupone el libre albedrío, la autoconciencia y el plan de vida". Por otra parte, añade Rafael de Asís, la identidad personal "presupone una idea de identidad humana, que es una suerte de universalización de las identidades personales: aquello que es común a todas ellas y que nos identifica como seres humanos". Por último, la identidad personal tampoco puede confundirse con la identidad jurídica (identidad pública del individuo como ciudadano) ni con la identidad digital (que se corresponde con nuestra imagen y reputación en el ámbito digital). Cfr., De Asís, 2022, 22-25.

la robótica avanzada, representan para la mejora de la calidad de la vida de las futuras generaciones.

2. UN DEBATE ÉTICO-JURÍDICO EN TORNO A LOS NEURO-IMPLANTES Y EL USO TERAPÉUTICO DE LA INTELIGENCIA ARTIFICIAL

La transformación digital está cambiando a un ritmo tan vertiginoso que, cuando apenas hemos empezado a familiarizarnos con el Internet de las cosas, ya se está anunciando un salto evolutivo de la tecnología en su afán por explorar y ampliar las fronteras sensoriales de la red. En efecto, con el Internet de los sentidos se pretende fusionar el mundo real y digital hasta el punto de hacerlos indistinguibles. El objetivo de un hombre conectado a la red permite imaginar un futuro en el que el *homo excelsior* (un cíborg resultado de la simbiosis entre la máquina y el humano) pueda desarrollarse neurológicamente y experimentar a través de las tecnologías digitales los cinco sentidos. A propósito de la conexión neuronal entre el ser humano y las Nuevas Tecnologías Digitales mediante el uso de implantes subdérmicos, neurotransmisores, interfaces y microchips cerebrales son referenciales, por ejemplo, los proyectos de ingeniería neuronal de Elon Musk (a través de la empresa Neuralink) o Mark Zuckerberg (mediante el Metaverso VR de realidad virtual)[3].

A la hora de determinar cuál es la capacidad y dónde se sitúan los límites de la inteligencia humana desde un punto de vista científico, ante todo hay que considerar que gran parte de nuestra actividad cerebral se dedica a recibir y procesar la información sensorial que tanto influye en nuestros actos y toma de decisiones. En este sentido, Kevin Warwick, uno de los mayores expertos mundiales en IA y cibernética (considerado por muchos como el primer cíborg de la historia desde que en 2002 conectó los nervios de su brazo a una mano biónica), advierte la limitada capacidad del pensamiento humano para percibir potencialmente señales que no son perceptibles para los seres humanos, pero sí para los robots inteligentes desarrollados con IA. Teniendo en cuenta la limitada capacidad de la mente humana, la mayoría de las aplicaciones actuales de los sensores no humanos consisten precisamente en convertir dichas señales extrasensoriales para los humanos en energía que éstos puedan percibir, como, por ejemplo, una imagen virtual de rayos X. Según la previsión de Warwick, el

[3] Martha J. Farah ha sido una de las primeras investigadoras en analizar las implicaciones éticas de la tecnología neuroquirúrgica, con especial énfasis en el empleo de la neurofarmacología mediante neurotransmisores para el tratamiento de enfermedades como el Alzheimer, el Trastorno por Déficit de Atención e Hiperactividad, y también fue de una de las primeras autoras en plantear los efectos ético-jurídicos que produciría la posibilidad de acordar judicialmente un tratamiento modificador de conductas en personas con comportamientos asociales (Farah, 2002, 1123-1129).

empleo de la amplia gama potencial de entradas sensoriales por parte de los sistemas de IA irá aumentando claramente su gama de capacidades conforme vaya transcurriendo el tiempo (Warwick, 2012, 146, 173-174).

Una prueba de que la línea de separación entre el hombre y la máquina se estrecha cada vez más la encontramos en el sistema de implante cerebral *Braingate*. Hasta ahora, las interfaces cerebro-ordenador se han utilizado con fines terapéuticos, para superar un problema médico/neurológico. Sin embargo, también existe la posibilidad de emplear esta tecnología para dotar a los individuos de habilidades que, en general, no poseen los seres humanos[4].

Al margen de las múltiples ventajas terapéuticas que ofrecen los neuroimplantes, y de los potenciales efectos benéficos de la aportación tecnológico-sensorial para la mejora de la memoria o el avance en la investigación sobre la comunicación mental, un individuo con implantes neuronales y conectado con la IA también podría disfrutar de la rápida y alta precisión en términos de "cálculo de números", podría acceder a una base de conocimientos de alta velocidad, casi infinita, en Internet, desarrollar una memoria precisa a largo plazo y aumentar su capacidad de detección.

Sin embargo, pese a estos buenos augurios respecto a los efectos beneficiosos que la aplicación de la ingeniería informática y de la cibernética supone para el sector sanitario, hay que considerar también cuál es la realidad y conocer los límites de la naturaleza humana en relación con estas buenas perspectivas sobre la introducción de las NN.TT. en la medicina, en general, y la neurología en particular. A este respecto, observa Warwick, desde un punto de vista técnico, los seres humanos sólo pueden visualizar y comprender el mundo que les rodea en términos de una percepción tridimensional limitada, mientras que los ordenadores son muy capaces de manejar cientos de dimensiones (Warwick, 2015, 5).

Por otro lado, es conveniente también conocer qué implicaciones ético-jurídicas puede tener el avance de la IA y la robótica en el ámbito de las libertades, los derechos y las obligaciones de los seres humanos (hasta el punto de que se ha abierto un debate doctrinal reciente en torno al reconocimiento de

[4] Según la explicación de Kevin Warwick del funcionamiento del implante cerebral *Braingate*, la actividad eléctrica de unas pocas neuronas monitorizadas por los electrodos de la matriz es decodificada en una señal para dirigir el movimiento del cursor. Esto permitió a un paciente que se sometió voluntariamente a esta prueba de monitorización neurológica posicionar un cursor en la pantalla de un ordenador, utilizando señales neuronales para su control, combinadas con información visual. La misma técnica se empleó posteriormente para poder realizar diversas operaciones con un brazo robótico a un paciente que sufría parálisis en uno de sus brazos (Warwick, 2015, 4).

una nueva clase de derechos humanos: los "neuroderechos")[5]. Hay dos proyectos de investigación dirigidos a crear una infraestructura de vanguardia en el campo de la neurociencia[6], la computación y la medicina relacionada con el cerebro: el primero es el *BRAIN Project* (acrónimo de Brain Research through Advancing Innovative Neurotechnologies), dirigido por el científico español Rafael Yuste y que fue financiado por la administración norteamericana en 2013; el segundo es el proyecto europeo *Human Brain Project*. Los dos proyectos coinciden en su propósito de "mapear o cartografiar" la actividad neuronal por medio de técnicas de neuroimagen para descifrar la interconexión neuronal del cerebro humano en un futuro próximo (Morente Parra, 2021, 265).

En un artículo publicado recientemente en la revista *Horizons,* bajo el título: "Its time for neurorights" (2021), sus autores -entre los que se encuentra precisamente Rafael Yuste- parten del convencimiento de que los avances tecnológicos que marcarán el tránsito del individuo hacia el universo de la singularidad no solo están redefiniendo ya la vida humana, sino que incluso están transformando el rol de los seres humanos en su vida social. En el ámbito de la ingeniería biomédica, la neurotecnología (conjunto de herramientas o métodos para potenciar y estimular la actividad cerebral) es el campo donde más profundamente se está constatando la alteración del significado de lo que, hasta ahora, hemos considerado esencialmente humano; no en balde, el cerebro es el órgano encargado de generar toda nuestra actividad mental y cognitiva (Yuste/Genser/Herrman, 2021, 154-155).

Sin duda, el potencial transformativo de la neurotecnología supone una mejora de las condiciones de vida a corto-medio plazo, y permite concebir la idea de un salto en la evolución de la especie humana más a largo plazo; por otra parte, el carácter transformativo de la naturaleza humana por parte de la neurotecnología ha generado un debate en torno a la necesidad de crear un marco jurídico específico, que sirva para reconocer y amparar un nuevo catálogo de derechos humanos que llevan la etiqueta de "neuroderechos"[7].

---

[5] La primera alusión a los neuroderechos la hicieron J. Sherrod Taylor, J. Anderson Harp y Tyron Elliot en un artículo sobre la creciente colaboración entre neuropsicólogos y neuroabogados titulado así precisamente: "Neuropsychologists and neurolawyers", en *Neuropsychology*, vol 5 (4), October 1991, pp. 293-305. Sin embargo, han sido Marcello Ienca y Roberto Andorno quienes, en puridad, se han referido expresamente al término "neuroderechos" en un artículo titulado: "A New Category of Human Rights: Neurorights" (2017). Disponible en http://blogs.biomedcentral.com/bmcblog/2017/04/26/new-category-human-rights-neurorights/. Última consulta: 28 de abril de 2022.

[6] La neurociencia adquirió carta de naturaleza en el Congreso de San Francisco titulado: "Neuroethics: Mapping the Field", celebrado entre los días 13 y 14 de mayo; cfr., Marcus 2002.

[7] En el apartado XXVI de la Carta de Derechos Digitales (que no tienen carácter normativo, pero que sí posee un objetivo prospectivo respecto a la aplicación e interpretación de los derechos en el

(...)

Es fácil imaginar las múltiples ventajas que ofrecen las neurotecnologías aplicadas a las ciencias de la salud. Pensemos, por ejemplo, en el interfaz cerebro-ordenador (cuyas siglas en inglés son BCI: *brain-computer interface*) un sistema de comunicación que monitorizan la actividad cerebral y permiten accionar el dispositivo de control de mecanismos que permiten interactuar a personas con discapacidades o enfermedades degenerativas que reducen o impiden su motricidad (De Asís, 2014, 35-36).

Ahora bien, si bien es cierto que hay un anverso en el desarrollo de la tecnología, por ejemplo, en su capacidad para tratar patologías neurológicas, no puede soslayarse que hay la neurotecnología presenta también un reverso, ya que puede ser útil para otros fines completamente espurios y lesivos de los derechos humanos, como sucede con en el control mental del enemigo en el ámbito militar, con la tortura a los prisioneros de guerra para la extracción de información, o con cualquiera de los otros supuestos en los que, según los teóricos del Derecho penal del enemigo (*Feindstrafrecht*), estuviera justificada la legalización del uso de la neurotecnología para injerir en la voluntad de quienes no merecieran ser tratados como personas, sino como enemigos de la sociedad (Jakobs /Polaino Orts, 2009).

Pero sin llegar siquiera a plantearnos escenarios tan extremos en la utilización de la neurociencia como los que se acaba de mencionar, el acceso a la información almacenada en el cerebro humano podría plantear dilemas ético-jurídicos también en el ámbito de las relaciones laborales; en este sentido, cabría preguntarse qué sucedería si un algoritmo de contratación discriminara a un posible empleado de una empresa porque interpretara mal sus datos cerebrales pues, a fin de cuentas, los algoritmos son capaces de desarrollar prejuicios que imitan a los que tenemos los seres humanos, como la raza o el género (Yuste/Genser/Herrman, 2021, 159).

En cualquiera de los casos anteriormente referidos se demuestra que la neurotecnología puede ser objeto de abuso intencionado o accidental por parte de quienes recurren a ella, ya sea con una finalidad terapéutica o

entorno digital del futuro inmediato) se enuncian los fines a los que se orientan los derechos digitales en el empleo de las neurotecnologías (fines que algunos consideran directamente como los cinco neuroderechos fundamentales): a) garantía del control de cada persona sobre su propia identidad; b) garantía de la autodeterminación individual, soberanía y libertad en la toma de decisiones; c) asegurar la confidencialidad y seguridad de los datos obtenidos o relativos a sus procesos cerebrales y el pleno dominio y disposición de los mismos; d) regular el uso de interfaces persona-máquina susceptibles de afectar a la integridad física o psíquica; e) asegurar que las decisiones y procesos basados en neurotecnologías no sean condicionadas por el suministro de datos, programas o informaciones incompletos, no deseados, desconocidos o sesgados. La información oficial sobre este documento puede consultarse en https://www.lamoncloa.gob.es/presidente/actividades/Documents/2021/140721-Carta_Derechos_Digitales_RedEs.pdf

malintencionada. En la era de la revolución tecnológica, marcada por la omnipotencia y la omnipresencia de la IA, no pueden darse por ciertos ni el derecho a la identidad personal (entendida como el conjunto de atributos y características que permiten individualizar a la persona en la sociedad), ni el libre albedrío, ni la privacidad mental, ni el acceso equitativo al neuropotenciamiento, ni la protección contra sesgos y discriminaciones ocasionadas por el uso erróneo o interesado de la neurociencia. Por eso, al hilo de la necesidad de proteger los derechos y las libertades de los ciudadanos ante el posible uso invasivo y perverso de las neurotecnologías, se ha abierto un debate en torno a la conveniencia de crear un marco jurídico para la salvaguarda de los neuroderechos. En este sentido, esta iniciativa neurocientífica iniciada por Rafael Yuste y la *Neurorights Foundation* ha tenido especial eco en Chile, hasta el punto de que ha dado lugar a la tramitación de una enmienda constitucional (Ley 21.383, D.O. 25-10-2021) para reformar el artículo 19,1 de la Constitución Política de Chile e implementar leyes para definir y delimitar las condiciones bajo las cuales podría realizarse el tratamiento de los datos cerebrales, y para redactar un proyecto de ley de neuroprotección de la identidad mental, a modo de reconocimiento como nuevo derecho humano del cerebro y su funcionalidad como núcleo del libre albedrío, pensamientos y emociones que caracterizan y diferencian a la especie humana (López Hernández, 2021, 95).

En todo caso, como se ha puesto de manifiesto en este nuevo proceso constituyente de Chile, las discusiones mantenidas a propósito de la aprobación de este proyecto de ley sobre los neuroderechos han servido para que se visibilice el argumentario de quienes, por una parte, consideran prioritario el reconocimiento de una nueva generación de derechos, es decir, una cuarta generación de derechos humanos, encuadrados en la categoría de los derechos digitales, y quienes, por otra parte, entienden que legislar en torno a un contexto tecnológico-científico resulta aún tan prematuro, especulativo e hipotético que sería contraproducente en términos jurídico-políticos, en la medida en que, con el reconocimiento de un catálogo tan reducido y específico de neuroderechos, se estaría contribuyendo a la inflación y la relativización de los derechos humanos que ya están consolidados, y que solo necesitarían una reformulación que los actualizase y adaptase al *momentum* de transformación digital que está experimentando la sociedad tecnológica y, particularmente, el mundo del Derecho.

Frente a las posiciones antagónicas mantenidas por los apocalípticos y los integrados de cara a las Nuevas Tecnologías Digitales, hay quienes apelan a la "responsabilidad tecnológica", entendida como una actitud reflexiva y crítica de los nuevos problemas que suscitan la ciencia y la tecnología, y ante los que ni la democracia, ni la ciencia, ni el Derecho, ni las Humanidades pueden

permanecer impasibles, sobre todo por su repercusión en el alcance y ejercicio de los derechos humanos (Pérez Luño, 2012, 42-43).

3. LA NUEVA GENERACIÓN DE DERECHOS DIGITALES Y EL RECONOCIMIENTO DE LOS NEURODERECHOS

Desde su origen y desarrollo a partir de la década de los ´90 del pasado siglo, Internet se ha convertido en la primera red de comunicación del mundo y, aunque son múltiples las múltiples ventajas y utilidades que nos ofrece en lo relativo al acceso a una ingente cantidad de datos e información, tampoco conviene soslayar la transformación que está experimentando el modelo de espacio digital y que, por motivos de ciberseguridad y de intereses del mercado global, no solo está modificando el carácter abierto, libre y neutral con el que fue creada Internet, sino que también está afectando a la privacidad y a la identidad de sus millones de usuarios (los social *Big Data* establecen patrones de conducta y realizan un perfil de sus millones de usuarios mediante la recopilación masiva no solo de sus datos personales, sino también de sus creencias y emociones). A este respecto, comenta Moisés Andrés Barrio que gran parte de nuestra vida cotidiana ha migrado hasta tal punto a Internet que se ha convertido en un medio representativo de nuestra cultura, mientras que nosotros, los usuarios, "hemos transformado nuestras identidades" (Andrés Barrio, 2021, 206).

Habitualmente hacemos mención a Internet de todas las cosas para referirnos al acceso a una cantidad de datos e información tan inconmensurables que suponen la puesta a disposición de los usuarios de unas fuentes ilimitadas de conocimiento sin precedentes en la historia. Sin embargo, la transformación digital también debiera servir para garantizar la mejora de la calidad de la democracia y el ejercicio de los derechos de los ciudadanos. En otras palabras, no basta con concebir Internet como un universo artificial por el que circulan millones de datos, sino también como un espacio en el que se nos garantiza la protección y el libre ejercicio de nuestros derechos en el ámbito digital.

A raíz de la repercusión de la revolución en el mundo del Derecho ha emergido una nueva generación de derechos cuyo objetivo principal consiste en la corrección de los problemas y perjuicios causados a la ciudadanía debidos a la falta de una regulación apropiada capaz de establecer un marco jurídico específico para el uso, el despliegue y el desarrollo las tecnologías digitales e Internet, la IA, la robótica y las tecnologías conexas; se trata de los derechos digitales, unos derechos asentados conceptualmente sobre

> un soporte virtual, no analógico, donde el cuerpo se volatiliza para dar paso a una estructura distinta de derechos que han de buscar la

> seguridad de la persona sobre el tratamiento de los datos y la arquitectura matemática de los algoritmos (Andrés Barrio, 2021, 209).

El artículo 18.4 de la Constitución española, inspirándose en el art. 35 de la Constitución portuguesa de 1976, supuso una novedad al establecer el límite legal al uso de la informática para garantizar el honor y la intimidad de los ciudadanos; a partir de este precepto constitucional se desarrollaría un cuerpo normativo y una importante línea jurisprudencial a propósito de la protección de datos. Sin embargo, la protección de datos no es suficiente ni agota todas las opciones para satisfacer el necesario establecimiento de un marco de garantía y protección efectivo de los derechos y las libertades de los ciudadanos en la era digital. A este propósito responde, precisamente, la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales, y más recientemente, la Carta de Derechos Digitales (CDD) que, pese a carecer de fuerza normativa, tiene el valor de servir de referencia para una futura ley reguladora de los derechos digitales. Entre los derechos reconocidos por la CDD se encuentran los derechos ante la IA y la neurociencia (lo cual podría suponer una vía abierta para el futuro reconocimiento de los neuroderechos).

A propósito del reconocimiento de los neuroderechos, sobre todo a partir de la convergencia del desarrollo de las neurotecnologías y de su vinculación directa de los cerebros humanos con la IA, Rafael Yuste y Sara Goering han expresado su preocupación porque el desarrollo de los dispositivos comercializados por las empresas neurotecnológicas en los mercados de consumo general se produzca de acuerdo con unos principios éticos, y según unos mínimos estándares de calidad y buena praxis que al implantarse no resulten invasivos y presenten el menor riesgo posible para las personas. En este sentido, en relación con la conexión entre el cerebro humano y las máquinas dotadas de IA, bien a través de neuroimplantes o de interfaces, estos autores (junto a otros miembros del Grupo Morningside)[8] plantean cuatro esferas de preocupación (*four concerns*) en las que se pone de manifiesto la necesidad de que el desarrollo y la aplicación de las nuevas neurotecnologías, como la estimulación cerebral profunda y la interfaz cerebro-computadora, se lleve a cabo conforme a los principios éticos de la neurotecnología y de la IA, de modo que se pueda garantizar el respeto y la preservación de la privacidad, la identidad, la agencia y la igualdad de las personas (Yuste/Goering, 2017, 159-163).

La primera preocupación de estos autores se debe a los efectos que la interacción entre la neurociencia y la IA pueden causar en la salvaguarda de la

[8] El Grupo Morningside está formado por neurocientíficos, neurotecnólogos, médicos, especialistas en ética e ingenieros de inteligencia artificial.

privacidad y el respeto al consentimiento de los pacientes que no deseen compartir sus datos neuronales. En este sentido, proponen que se regule la venta, la transferencia comercial y el uso de datos neuronales (una regulación parecida a la *US National Organ Transplant Act* de 1984). Otra medida de protección de la privacidad del usuario de las neurotecnologías podría ser la aplicación de técnicas basadas *blockchain* y *smart contracts* que propician, sin la intermediación de una autoridad centralizada, una información transparente sobre cómo se están administrando los datos de la actividad neuronal de los individuos.

El segundo motivo de inquietud de Rafael Yuste y Sara Goering plantea la hipótesis de que las neurotecnologías y la IA lleguen a alterar el sentido de la identidad y agencia racional de las personas, pudiendo incluso subvertir la propia naturaleza del yo y la responsabilidad moral y jurídica del individuo. En efecto, de confirmarse la pérdida de nuestro sentido de la agencia y de la identidad (por ejemplo, a través de dispositivos de control neuronal que monitoricen a distancia el pensamiento o mediante la interconexión de varios cerebros que trabajen a la vez en colaboración, los individuos podrían terminar comportándose de una forma ajena a su verdadera personalidad, hasta el punto de que ni ellos mismos podrían reconocerse en sus actos. Como posible solución a esta segunda preocupación, Yuste y Goering proponen la inclusión de cláusulas protectoras de los neuroderechos en los tratados internacionales, y la creación de una Convención internacional para definir las acciones prohibidas relacionadas con la neurotecnología y la IA, similar a las prohibiciones enumeradas en la Convención Internacional para la Protección de Todas las Personas contra las Desapariciones Forzadas (que entró en vigor el 23 de diciembre de 2010).

La tercera razón de desasosiego de los autores vinculados al Grupo Morningside tiene que ver con el aumento de capacidad cognitiva y el neuropotenciamiento que actualmente es una de las puntas de lanza del transhumanismo tecnológico. En este sentido, Laurent Alexandre, prestigioso médico y neurobiólogo transhumanista francés, ha advertido que la única salida que le queda a la humanidad ante el inevitable advenimiento de la singularidad tecnológica es "coevolucionar" con las máquinas y potenciar tecnológicamente el cerebro humano para adaptarlo a la IA fuerte que, según su pronóstico, determinará el futuro posthumano (Alexandre, 2018, 291 y ss). Ante este panorama, Yuste y Goering consideran probable que el nivel de presión para adoptar neurotecnologías potenciadoras llegue a tal grado que termine cambiando los usos y las reglas sociales desde un punto de vista ético-político, e incluso que genere problemas de acceso equitativo y nuevas formas de discriminación (fractura tecnológica). Por eso, ambos autores proponen establecer límites ético-jurídicos al desarrollo de las neurotecnologías y definir

los contextos en los que se pueden aplicar (como sucede, por ejemplo, con la edición genética realizada en seres humanos), pero sin llegar a imponer prohibiciones absolutas a ciertas tecnologías (como las que estimulan y potencian al cerebro humano) que solo servirían para empujarlas a la zona oscura de la clandestinidad.

El cuarto motivo de preocupación compartido por Rafael Yuste y Sara Goering es el de los sesgos o prejuicios (*bias*) que tan influyentes resultan, por ejemplo, en los procesos selectivos o resolutivos en los que se recopilan infinidad de datos personales de trabajadores mediante técnicas de *data mining* y de discriminación algorítmica que se ponen al servicio de los responsables de optimizar los recursos humanos de una empresa (*workforce analytics*). A este respecto, conviene tener en cuenta que, como advierte Serena Vantin, el uso de instrumentos algorítmicos en el ámbito laboral y empresarial no se limita solo a las técnicas de *workforce analytics*, sino que también se extiende a la digitalización de los procesos productivos, a los servicios de *gig economy* (una fórmula de contratación online y absolutamente flexible para el empleador y el empleado que se presenta como alternativa al modelo de contrato fijo tradicional), a las nuevas técnicas de vigilancia de los empleados por parte de los empresarios en horario de trabajo, etc. (Vantin, 2021, 96-97).

Como vemos, el enorme potencial que ofrece el uso de los algoritmos para facilitar el acceso de la ciudadanía a la Administración pública más transparente y eficaz, para garantizar nuestra seguridad y el ejercicio de nuestros derechos, o para impulsar la modernización de las empresas, tiene también un reverso oscuro en el que los riesgos de discriminación digital tanto en la red, como en los sistemas de IA, robótica y tecnologías anexas (Pietropaoli, 2019, 379-400). Por otra parte, los sesgos discriminatorios, los prejuicios contrarios a la dignidad y al derecho a la igualdad y los errores algorítmicos no perjudican uniformemente a toda la población, sino que suelen afectar especialmente a los grupos más vulnerables y a los individuos más desfavorecidos dentro de la sociedad (Vantin, 2021, 96).

A propósito de los sesgos discriminatorios, Yuste y Goering recomiendan la participación de los usuarios probables -y especialmente de los que se encuentren marginados- en el diseño de algoritmos y dispositivos desde su primera fase de desarrollo tecnológico precisamente para evitar situaciones de sesgos discriminatorios en los sistemas de toma de decisión algorítmica (*algorithmic decision making*). En los últimos años, algunos estudiosos de los procesos de toma de decisión algorítmica están investigando sobre el modo de revertir el uso de algoritmos selectivos en un sentido equitativo, y de acuerdo

con la garantía de transparencia contemplada en la estrategia digital europea[9]: me refiero a los *Critical Data Studies* (Lettieri, 2020, 54-55).

Una buena síntesis del actual debate doctrinal en torno a la necesidad de construir una teoría de neuroderechos como derechos humanos nos la proporciona Rafael de Asís en su libro *Derechos y tecnologías* (2022a). Según se pone de relieve en este estudio monográfico, hay una incipiente línea doctrinal iberoamericana en la que se propugna el reconocimiento de una nueva generación de derechos humanos, a partir de la proclamación de los *cinco neuroderechos*[10] propuestos por Rafael Yuste, Jared Genser y Stephanie Herrmann (2021, 160-161).

En este sentido, una postura representativa de esta doctrina favorable al reconocimiento de la vertiente ético-jurídica de los neuroderechos y a su incorporación intrasistemática en el ordenamiento jurídico, mediante su positivación y reconocimiento como pertenecientes a una cuarta generación de derechos humanos, es la mantenida por Enrique Cáceres Nieto, Javier Díaz García y Emilio García García (2021, 79-80). Esta línea doctrinal favorable al reconocimiento de los neuroderechos también cuenta con un marco institucional de *softlaw* regional: la Declaración del Comité Interamericano sobre "Neurociencia, Neurotecnologías y Derechos Humanos: Nuevos Desafíos Jurídicos para las Américas"[11], y sigue la misma estela trazada anteriormente

[9] Dentro del marco de las instituciones europeas existen algunos estudios sobre el procedimiento de toma de decisiones algorítmicas; véanse, por ejemplo, a este respecto: "Understanding Algorithmic Decision-making. Opportunities and Challenges", 2019, disponible en: https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS_STU(2019)624261_EN.pdf; "A Governance ramework for Algorithmic Accountability and Transparency, 2019, disponible en: https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf; sobre la estrategia digital "Shaping Europe´s Digital Future", 2020, disponible en:, https://ec.europa.eu/info/sites/default/files/communication-shaping-europes-digital-future-feb2020_en_4.pdf

[10] Los cinco neuroderechos propuestos por Yuste, Genser y Hermann son: 1.- el derecho a la identidad, o la capacidad de controlar nuestra integridad física y mental ; 2.- el derecho a la libertad de pensamiento y al libre albedrío para decidir cómo actuar; 3.- el derecho a la privacidad mental, o la protección de nuestro pensamiento contra la divulgación; 4.- el derecho a un acceso justo para el aumento del potencial de la mente, es decir, la capacidad de garantizar que los beneficios de las mejoras de la capacidad sensorial y mental a través de la neurotecnología se distribuyan de forma justa en la población; y 5.- el derecho a protección contra los sesgos algorítmicos, o la garantía de que las tecnologías no introduzcan prejuicios.

[11] Esta Declaración se aprobó tras la reunión mantenida por el Comité Jurídico Interamericano entre los días 2-11 de agosto de 2021, dentro del 99º periodo ordinario de sesiones, y se publicó el 4 de agosto de ese mismo año. El texto está disponible en la siguiente dirección: https://kamanau.org/wp-content/uploads/2021/08/Neuro-derechos-doc-641-rev-1-esp-DN-ROA.pdf Última consulta: 28 de abril de 2022.

por una doctrina favorable a la aprobación de una Declaración Universal de los Neuroderechos Humanos (Sommaggio/Mazzocca/Gerola/Ferro, 2017, 27-45).

Otros autores son más remisos a la propuesta de ampliar el catálogo de derechos humanos, alegando que con la profusión de los mismos se generan problemas de indeterminación e incoherencia en su fundamentación, además de un posible debilitamiento de su eficacia al solaparse con derechos humanos de generaciones anteriores. En este sentido, resulta elocuente la posición de Francisco Laporta contraria a rebajar el rigor en el proceso de reconocimiento de nuevos derechos humanos (como los relacionados, precisamente, con las Nuevas Tecnologías); a este respecto señala este autor:

> Me parece razonable suponer que cuanto más se multiplique la nómina de los derechos humanos menos fuerza tendrán como exigencia, y cuanto más fuerza moral o jurídica se les suponga más limitada ha de ser la lista de derechos que la justifiquen adecuadamente (Laporta, 1987, 23).

A propósito de la superposición de los neuroderechos en relación a los derechos y las libertades consagrados en la Declaración Universal de los Derechos Humanos (DUDH), hay autores que sostienen que su reconocimiento no se justifica si los bienes jurídicos que pretenden garantizar los neuroderechos: la intimidad, la privacidad, la libertad, la dignidad humana y el acceso equitativo a los recursos científicos, ya han sido reconocidos y garantizados antes tanto en la DUDH, como en los pactos y convenios internacionales posteriores (Morente Parra, 2021, 273; en sentido análogo, Borbón/Borbón/Laverde, 2020, 146).

En una posición intermedia dentro de este debate sobre la oportunidad del reconocimiento de los neuroderechos se mantiene Rafael de Asís, a quien le produce perplejidad el hecho de que en el proceso de incorporación de las NN.TT. en el ámbito educativo se esté dejando de lado e incluso rechazando la educación en derechos humanos (la necesaria formación tecnológica de nuestros estudiantes no solo no es incompatible, sino complementaria con la formación en Humanidades y la transmisión de la cultura de los derechos humanos[12]. En cualquier caso, concluye este autor, la aplicación a las cuestiones

[12] A propósito de la importancia de la educación en los derechos humanos, Manuel Atienza señala que aunque el conocimiento y la educación no bastan para terminar por sí solos con el mal en el mundo, sin embargo, resultan imprescindibles: "la lectura de los textos que recogen las declaraciones de derechos humanos, la reflexión en torno a los diversos problemas que plantean y, en general, la incorporación de esa materia (teórica y práctica) a los currículos de las escuelas y de las universidades y su presencia en los foros de discusión pública no van a lograr probablemente un significativo efecto de persuasión en los grandes poderes (en parte públicos pero, sobre todo, privados) de este mundo, que son los principales responsables de que esos

(...)

sociales de las NN.TT., en general, y de las neurotecnologías, en particular, "es una realidad que conviene afrontar" (De Asís, 2022a, 148-152; 2022b, 148).

Por consiguiente, más que de una repetición de derechos con distinta etiqueta, se trataría de hacer un ejercicio de concreción dentro de la fase de especialización de los derechos humanos que, si son contemplados desde una perspectiva histórica, es decir, en su dimensión diacrónica o evolutiva a lo largo del tiempo, de acuerdo con la tesis de la mutación histórica de los derechos humanos (*Wandel der Grundrechte*), no deberían convertirse en conceptos fosilizados dentro de un catálogo de derechos y libertades intemporales incorporados a una lista con *numerus clausus*. Como ya advirtiera en la década de los años 80 del pasado siglo Antonio E. Pérez Luño, al hilo de la concepción generacional de los derechos humanos, los derechos y libertades de nueva generación se presentan como una respuesta al proceso de erosión y degradación que aqueja a los derechos fundamentales ante determinados usos de las NN.TT. (problema al que la doctrina anglosajona se refiere con el término *liberties' pollution*). Las consideraciones que hacía Pérez Luño, a propósito de la "sociedad de la información" y del interés prioritario que tenía la regulación jurídica del uso de la informática, bien podrían ampliarse hoy a la sociedad tecnológica y a la necesidad de establecer un marco jurídico en torno al uso de las nuevas tecnologías NBIC y el desarrollo de la IA y la robótica avanzada (Pérez Luño, 1987, 58).

En el siguiente epígrafe me ocuparé especialmente de la identidad personal (conjunto de rasgos específicos que hacen única a una persona) ante los retos que le depara el metauniverso digital. El concepto de identidad adquiere pleno sentido cuando se complementa con nuevos derechos y libertades, como el derecho al libre desarrollo de la personalidad, la integridad mental y la libertad cognitiva (la libertad de controlar la propia conciencia); por cierto, una libertad, esta última, estrechamente vinculada con la clásica la libertad de pensamiento (Sententia, 2004, 222), aunque adaptada a las circunstancias del siglo XXI, y que ha sido definida por Richard G. Boire como "la quintaesencia de la libertad" (*the quintessence of freedom*) (Boire, 2001, 8).

derechos no estén garantizados para una inmensa mayoría de los habitantes del planeta. Pero todo ello sí que puede contribuir a que mucha gente adquiera conciencia de cuáles son los derechos que legítimamente puede reivindicar (y de los deberes que debe asumir) y de cuáles son las causas que impiden que los mismos puedan realizarse. Y si esa conciencia moral esclarecida se generalizase suficientemente, sería muy probable que se convirtiera también en una fuerza socialmente irresistible" (Atienza, 2020, 152).

4. CUANDO LA PERSONA SE CONVIERTE EN UN AVATAR: EL *NOVO HOMO LUDENS* EN EL METAUNIVERSO DE INTERNET

Se ha hecho anteriormente referencia a las esperanzas abiertas por las nuevas neurotecnologías, como la estimulación cerebral profunda (DBS) y el interfaz cerebro-computadora (BCI), en la prevención, el tratamiento y la curación de enfermedades como el párkinson, la epilepsia, el ELA o el trastorno obsesivo consultivo (TOC), pero tampoco deben soslayarse los efectos contraproducentes que esos dispositivos pueden tener en la identidad, la autenticidad y la autonomía de la persona. Para bien o para mal, lo cierto es que estos dispositivos son capaces de interferir en la autoconciencia y alterar la agencia de los individuos en los que se implantan (Goering/Brown/Klein, 2021).

Al invocar el señorío de nuestra mente como un derecho innato y no adquirido, también estamos apelando a la inalienabilidad de nuestra identidad personal, a la inviolabilidad de nuestra integridad física y mental, a la preservación de nuestra autenticidad, a la capacidad de decidir libremente nuestra actuación (facultad que también se conoce como "control agencial"), y a la autonomía de nuestra voluntad (Bublitz, 2013, 7-11).

El problema aparece cuando el individuo pierde inconscientemente el control de su autonomía debido a factores o agentes externos que interfieren sus facultades mentales, nublan su juicio y dirigen su conducta (Bublitz/Merkel, 2009, 371). Esta manipulación inadvertida del individuo agente rompería la continuidad psicológica mediante la introducción de un hiato entre sus preferencias actuales del agente y las que tenía arraigadas en su personalidad cuando era un sujeto psicológicamente autónomo hasta que se produjo dicha injerencia desde el exterior (Mele, 1995, 187; Hagi, 1998, 108 ss.; Kapitan, 2000, 81-104).

La línea de separación entre el hábito y la dependencia del *phono sapiens* (el *novo homo ludens*) respecto a los dispositivos electrónicos y digitales que éste utiliza en su vida cotidiana es tan tenue a veces que no resulta fácil diferenciarla, y la aparente libertad de elección de usar la yema de los dedos sobre la superficie de la pantalla de su ordenador portátil, teléfono móvil o tablet no es más que "una selección consumista" (Han, 2021, 24).

En esos intervalos diarios de ausencia del hombre de su realidad, en ese sustraerse de sus circunstancias y de las cosas del mundo real, y en su poder de retirada virtual y provisoriamente del mundo y meterse dentro de sí, se produce un fenómeno característico del ser humano del que carecen otros animales: el "ensimismamiento" (Ortega, 2006a, 536). Para Ortega y Gasset este acto de ensimismamiento, esta retirada estratégica a sí mismo, es un privilegio con el que el hombre consigue de liberarse transitoriamente de las cosas

precisamente a través del dominio de la técnica, cuya misión inicial consiste, precisamente, en "dar franquía al hombre para poder vacar a ser sí mismo", es decir, crear un espacio *extranatural* de ocio (*otium*) que se abre al hombre para que éste pueda ocuparse de algo más que de cubrir sus necesidades más elementales, como imaginar, inventar y crear, tanto en el campo de las ciencias como en el de las artes (Ortega, 2006b, 574-575).

Al igual que Sartori denunciaba en *Homo videns*, la influencia que los medios de comunicación, y de modo especial la televisión, ejercía sobre las masas, un cuarto de siglo después nos encontramos con una situación parecida de enajenación por parte del *novo homo ludens,* con la única salvedad de que ahora son las neotecnologías NBIC bajo el dominio de las *Big Tech*, que se hallan en el contorno de ese individuo, quienes dirigen y controlan sus hábitos vitales e incluso su voluntad como si fuese una marioneta movida por los metadatos y los algoritmos que configuran el inescrutable universo de Internet. Esta situación aproxima al hombre a la alteración característica de la vida animal y le aleja de la autoconsciencia y del ensimismamiento humanos. Ortega lo explica primorosamente en *Ensimismamiento y alteración* (1939):

> Decir, pues, que el animal no vive desde *sí mismo* sino desde *lo otro*, traído y llevado y tiranizado por *lo otro*, equivale a decir que el animal vive siempre alterado, enajenado, que su vida es constitutiva *alteración* (Ortega, 2006b, 535).

A propósito de la alienación y la alteración del hombre contemporáneo, advierte Sartori que la nuestra es una época extraordinaria en la que quienes aún conserven la capacidad crítica de los seres pensantes tienen el deber de denunciar la irresponsabilidad e inconsciencia de las cada vez mayores

> legiones de vendedores de humo que olvidan que vivimos y viviremos no es "naturaleza" (una cosa dada que está ahí para siempre), sino que es de cabo a rabo un producto artificial construido por el *homo sapiens*. ¿Se podrá mantener sin su apoyo? No, seguramente no. Y si hacemos caso a los falsos profetas que nos están bombardeando con sus multi-mensajes, llegaremos rápidamente a un mundo virtual que se pone patas arriba en una "catástrofe real" (Sartori, 2018, 197).

La expansión del espacio digital más allá de los límites imaginables por Giovanni Sartori hace más de veinte años no solo ha difuminado la tenue línea de separación entre la naturaleza y la realidad virtual que ya entonces discernía con dificultades el filósofo y politólogo florentino, sino que en algunos ámbitos está absorbiendo incluso a la identidad humana, me refiero al mundo virtual del metaverso, de la realidad en tres dimensiones (3D) aumentada capacidad

5G, la Inteligencia Artificial y el inminente desarrollo del Internet de los sentidos que pretende usar el cerebro como interfaz, modular con microimplantes el mundo sonoro que nos rodea, personalizar el sabor de los alimentos, o incluso recrear (o crear *ex novo*) aromas y otros sentidos digitales como el tacto.

La realidad humana parece haberse visto superada por la ficción mecánica del mundo digital cuando ya cabe concebir la amistad y hasta el enamoramiento virtual con una máquina desarrollada mediante IA, o con un personaje de fantasía o avatar diseñado virtualmente; esta es, por cierto, una tendencia creciente en Japón, como demuestra el curioso, o más bien bizarro caso del Sr. Akihito Kondo, casado con una célebre cantante manga llamada Hatsune Miku con millones de fans, algo que no tendría nada de peculiaridad salvo por el hecho de que se trata de un holograma que tiene "existencia" virtual como *Vocaloid* (o cantante virtual) en un dispositivo denominado Gatebox que no solo le da vida como si fuera un tamagotchi sentimental, sino que ha llegado a formalizar el matrimonio entre un hombre y un holograma en un documento sin validez jurídica. A propósito de esta confusión entre la realidad humana y la ficción digital, algún estudio reciente sobre los efectos ético-jurídicos de la disociación humana ha advertido que cuando la identidad humana trata de conectar con un fetiche cibernético entonces es señal de que inexorablemente existe una propensión a descender al terreno de lo virtual y a olvidar la consciencia de la identidad humana en el continente digital (algo parecido a entrar en un trance que nos sumergiera en un sueño digital inducido tecnológicamente) (Curcio, 2020, 56).

Al margen del espejismo que produce en la psique humana la realidad virtual, y de la interacción entre la figura humana perfilada y reproducida en el continente digital recreado por el metaverso, los interfaces y los videojuegos 3D, lo cierto es que los humanos y las máquinas no son ontológicamente iguales, ni pertenecen a la misma categoría: los hologramas son imágenes tridimensionales configuradas con números y algoritmos, mientras que los seres humanos estamos hechos de carne y hueso, *ratio et emotio* (Illich, 1992; Curcio, 2020, 56).

La cada vez más tenue línea de separación entre el mundo natural-real y el universo digital-artificial nos previene del riesgo de minusvalorar la necesidad de preservar la identidad humana. Por eso, retomando la diatriba sobre la oportunidad de reconocer o no los neuroderechos, parece razonable al menos plantearse si, tal vez, ante la pérdida de conciencia de la realidad por parte del *novo homo ludens*, no tendría sentido proteger al menos el primero de esos neuroderechos, es decir, el derecho a la identidad, o la capacidad de controlar nuestra integridad física y mental.

De acuerdo con el criterio de la perspectiva generacional de los derechos humanos, cuyo catálogo no está formado por un elenco cerrado de derechos y libertades, sino por una lista abierta a los cambios y problemas más acuciantes que afectan al hombre contemporáneo en la era de las nuevas tecnologías (Vašák, 1990, 297), cabría sumar una cuarta generación en la que estaría integrado precisamente el derecho a la identidad humana. De la misma forma que la primera generación correspondería a los derechos y libertades individuales; la segunda, a los derechos económicos, sociales y culturales; y la tercera a las garantías jurídicas-subjetivas fundamentales propias de la era tecnológica; y de igual modo que cada una de esas generaciones se correspondería con los valores-guía de la libertad, la igualdad y la solidaridad, respectivamente (Pérez Luño, 2006, 232; 2018, 692-702), podríamos concluir que la cuarta generación se referiría a aquellos derechos y libertades protectores de la condición humana frente a los embates del transhumanismo tecnológico, y cuyo principio guía sería precisamente la dignidad humana.

La cuarta generación de derechos humanos se justifica en un escenario virtual, determinado por la IA, e integrado por recreaciones virtuales que provocan en el internauta la alucinación de interactuar con no-cosas que ni *son* ni *están* en la realidad física, pero que influyen cada vez más en su rutina diaria e incluso en su conducta. La actuación del individuo en el entorno digital, por más que sea artificial, tiene consecuencias jurídicas que le vinculan; por ejemplo, la tecnología *blockchain* ha posibilitado la realización de contratos inteligentes (*smart contracts*) escritos en lenguaje virtual, cuya ejecución es autónoma y automática, a partir de unos parámetros programados, y que ofrecen unas condiciones de seguridad, transparencia y confianza a las partes contratantes superiores a las de los contratos tradicionales en los que el riesgo de que haya malentendidos, falsificaciones o alteraciones es mayor. Esta misma vinculatoriedad de los contratos y negocios jurídicos suscritos en el espacio digital se constata en el creciente campo de las criptomonedas (no exentas del riesgo de la especulación y de la consiguiente devaluación) y de los NFT (activos digitales no fungibles), creados con *tokens* criptográficos al igual que las criptomonedas para determinar su autoría y singularidad, y que han revolucionado el mercado del arte digital hasta el punto de que en el último año se han multiplicado exponencialmente sus ventas e incluso su valor (en 2021, Jack Dorsey, cofundador de Twitter, vendió el primer tuit de la historia de su compañía por 2.95 millones de dólares, y el artista digital Beeple vendió un NFT en Christie´s por 69 millones de dólares).

El metaverso no es un concepto reciente, como se recordará, a principios del presente siglo se lanzó *Second Life*, una plataforma multimedia en línea en la que los usuarios creaban un avatar y construían una segunda vida digital. Con el transcurso del tiempo, este metaverso original diseñado por la compañía

tecnológica Linden Lab se convirtió en un arquetipo de metaverso que serviría como referencia a otros metaversos desarrollados posteriormente en la web 2.0 y en la web 3.0. En resumidas cuentas, el metaverso no se consiste en una experiencia unitaria en un espacio digital compacto, sino en la migración de la experiencia humana desde el mundo físico hasta numerosos mundos virtuales en los que, como sostienen los autores de un estudio reciente sobre el futuro marco jurídico del metaverso, la tecnología tiene la oportunidad de llevar contenido a esos mundos de maneras nunca antes imaginadas y, con ello, problemas y desafíos legales nunca antes contemplados (Ara/Radcliffe/Fluhr/Imp, 2022).

La progresiva implantación del metaverso (en el ámbito de la diversión, del comercio, de la salud y de la educación) ha generado una serie de supuestos y novedades desconocidos hasta ahora en nuestra experiencia jurídica. Es cierto que, en algunos casos, se podrían ajustar algunas leyes existentes para la regulación de cuestiones novedosas planteadas por la irrupción de las Nuevas Tecnologías; sin embargo, si se considera la inconmensurabilidad del espacio abierto en el que se expande el metaverso, cabe deducir que la adaptación legal y jurisprudencial a esa nueva realidad virtual que es jurídicamente vinculante no será fácil, en la medida en que las leyes existentes resultan ya insuficientes para regular los problemas causados en el espacio digital por un metaverso que ha roto las costuras de los sistemas jurídicos existentes.

En efecto, como señalan los autores del artículo sobre la regulación del metaverso anteriormente citado, el alcance de todas las leyes y regulaciones que podrían estar implicadas en un metaverso es prácticamente ilimitado y puede generar innumerables problemas legales. Así, por ejemplo, en materia de propiedad intelectual, la creación de nuevos tipos de NFT ha causado no pocas controversias y consultas legales respecto al alcance del derecho a utilizar el contenido en poder del propietario del NFT (en la praxis judicial más reciente la mayoría de las reclamaciones relativas al contenido del metaverso afectan a los derechos de autor, marcas comerciales y derechos de publicidad). Por otra parte, el uso y la explotación de los derechos de propiedad intelectual previamente licenciados o adquiridos en el metaverso plantean cuestiones novedosas para los licenciatarios y adquirentes en torno a la amplitud y el alcance de los derechos que han obtenido en virtud de acuerdos que pueden haber precedido durante mucho tiempo a Internet, y en menor medida al metaverso.

La problemática de los proyectos metaversos se extiende también a otras áreas legales, como, por ejemplo, las de la intimidad y la ciberseguridad.

En relación con la garantía de la privacidad en el proceso de recopilación, uso y transmisión de datos personales, los metaversos tienen capacidad para

recopilar una información muy diversa que puede ir desde la información básica de identificación hasta recabar datos sobre el movimiento y las actividades del usuario en el metaverso. A este respecto, por un lado, se va evidenciando cada vez más la necesidad de aprobar una legislación dedicada precisamente a la protección de la intimidad en el ámbito del metaverso e incluso, junto a la oportunidad de contar con una jurisdicción especializada en Derecho digital e IA jurídica; por otro lado, también los creadores y desarrolladores de los proyectos metaverso deberían considerar la implementación de medidas que aseguren el cumplimiento de los requisitos legales de privacidad y la observancia de unos mínimos estándares ético-jurídicos en los contenidos de los metaversos (Moore, 2021).

Respecto a la cuestión de la ciberseguridad, los proyectos metaversos plantean también problemas y cuestiones novedosas a las compañías tecnológicas que los crean y desarrollan, sobre todo de cara a asegurar la protección de sus sistemas de información y procesamiento de datos personales de sus usuarios ante un eventual ciberataque (Brighi, 2021, 133-147).

En definitiva, aunque el metaverso se encuentre todavía en una fase inicial de implantación tecnológica, a medida que vaya evolucionando y expandiéndose su uso, tanto a nivel profesional como doméstico, es presumible que también se incrementarán el número de incidencias y reclamaciones entre los usuarios; precisamente por eso se hará cada vez más evidente la necesidad de establecer un marco regulatorio del metaverso para tratar de anticipar -en la medida de lo posible- respuestas legales a los nuevos problemas legales que presente el metaverso (Ara/Radcliffe/Fluhr/Imp, 2022).

5. CONCLUSIÓN

El impacto que sobre los derechos y libertades produce la revolución tecnológica 4.0 desborda el ámbito de las tres generaciones anteriores de derechos y libertades, porque ahora el hombre contemporáneo no está solo ante la técnica, sino que coexiste en el espacio digital con otras entidades y otro tipo de inteligencias que no son estrictamente humanas, sino transhumanas y/o artificiales. El escenario posthumano que se abre ante nosotros es, por ende, más complejo e incierto que aquél que respondía al paradigma humanista y al canon antropocéntrico en el que fue posible alumbrar una fase de esplendor para el proyecto humanista de la modernidad, y que Norberto Bobbio definió como "el tiempo de los derechos" (*l´età dei diritti*). Este nuevo escenario posthumano nos sitúa frente grandes cuestiones y retos como la identidad humana y el metaverso, el status jurídico de los robots, la regulación del espacio digital, la fundamentación de una ética de la IA, la metamorfosis del Derecho y la Justicia, en suma, nos coloca ante un mundo en el que, como advierte Luciano Floridi, la humanidad intentará transformar un entorno artificial hostil

en una *infosfera* adaptada tecnológicamente en la que ésta perderá progresivamente su protagonismo. En efecto, señala este autor, en este nuevo habitat digital compartiremos espacio virtual "no solo con otras fuerzas y fuentes de acción natural, animal y social, sino también y sobre todo con agentes artificiales" (Floridi, 2022, 58).

Las revoluciones, escribía Antonio Gramsci, representan una forma de *hegemonía cultural.* La revolución digital no solo ha conseguido imponerse a las sociedades modernas como un universo cultural de referencia, sino también como una idea dominante que todos hemos interiorizado y hecho nuestra de algún modo. La revolución 4.0, que según Floridi se remonta a Alan Turing, nos coloca en un contexto de metamorfosis del mundo en donde se halla en juego la conservación de la esencia humana ante el horizonte de la singularidad tecnológica, en el cual "la inteligencia ya no es solo una prerrogativa humana sino también artificial y digital" (Balbi, 2022, 42).

6. BIBLIOGRAFÍA

Alexandre, Laurent (2018), *La guerra delle intelligenze. Intelligenza artificiale contro intelligenza umana*, trad. it., N. Nappi, EDT, Torino.

Andrés Barrio, Moisés (2021), "Génesis y desarrollo de los derechos digitales", en: *Revista de las Cortes Generales*, 110, pp. 197-233.

Ara, Tom K.- Radcliffe, Marcos F.- Fluhr, Miguel- Imp, Katherine (2022), "Exploring the Metaverse. What Laws will apply?". *DLA Piper*-Chambers TMT. February 22th 2022. Disponible en: https://www.dlapiper.com/en/latinamerica/insights/publications/2022/02/exploring-the-metaverse/

Arendt, Hannah (1959) *The Human Condition. A Study of the Central Dilemmas Facing Modern Man* (1958), Doubleday Anchor Books, Garden City (New York). trad. esp., R. Gil Novales, Paidós, Barcelona.

Atienza, Manuel (2020), *Una apología del Derecho y otros ensayos*, Trotta, Madrid.

Balbi, Gabriele (2022), *L´ultima ideologia. Breve storia della rivoluzione digitale*, Editori Laterza, Bari-Roma.

Bergson, Henri (1973), *La evolución creadora*, trad. esp., M. L. Pérez Torres, Espasa-Calpe, Madrid.

Boire, Richard G. (2001). "On cognitive liberty III", en: *Journal of Cognitive Liberties,* 2, 7-22.

Borbón Rodríguez, Diego A., Luisa Borbón Rodríguez, Jennifer Laverde Pinzón (2020), "Análisis crítico de los NeuroDerechos Humanos al libre albedrío y al acceso equitativo a tecnologías de mejora", en: *Ius et Scientia*, 6, 2, 135-161.

Brighi, Raffaella (2021), "Cybersecurity. Dimensione publica e privata della sicurezza dei dati", en: Casadei, Thomas, Stefano Pietropaoli, *Diritto e tecnologie informatiche. Questioni di informática giuridica, prospettive istituzionali e sfide sociali*, Milano, Wolster Kluwer, 135-147.

Bublitz, Jan Christoph (2013), "My Mind is Mine!? Cognitive Liberty as a Legal Concept", en: Hildt, E, A. Francke (eds.), *Cognitive Enhancement*, Berlin, Springer, 233-264.

Bublitz, Jan Christoph, Reinhard Merkel (2009), "Autonomy and Authenticity of Enhanced Personality Traits", en: *Bioethics* 25, 6, 360-374.

Cáceres Nieto, Enrique, Javier Díaz García, Emilio García García (2021), "Neuroética y neuroderechos", en: *Revista del Posgrado en Derecho de la UNAM* 15, julio-diciembre, 37-86.

Curcio, Renato (2020), *Identità cibernetiche. Dissociazioni indotte, contesti obbliganti e comandi furtivi*, Edizioni Sensibili alle foglie, Roma.

De Asís Roig, Rafael (2014), *Una mirada a la robótica desde los derechos humanos*, Instituto de Derechos Humanos "Bartolomé de Las Casas" de la Universidad Carlos III de Madrid, Dykinson, Madrid.

— (2022a). *Derechos y tecnologías*, Dykinson-Departamento de Derecho Internacional Público, Eclesiástico y Filosofía del Derecho de la Universidad Carlos III de Madrid.

— (2022b) "Sobre la propuesta de los neuroderechos", en: *Derechos y libertades. Revista de Filosofía del Derecho y derechos humanos*, 47, en imprenta.

Farah, Martha J (2002), "Emerging Ethical Issues in Neuroscience", en: *Nature Neuroscience* 5, 1123-1129.

Floridi, Luciano (2022), *Ética dell´intelligenza artificiale. Sviluppi, opportunità, sfide*, edizione italiana a cura di M. Durante, Raffaello Cortina Editore, Milano.

Goering, Sara - Brown, Timothy - Klein, Eran (2021), "Neurotechnology Ethics and Rational Agency", en: *Philos Compass*, 10. Disponible en: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8443241/

Hagi, Ishtiyaque (1998), *Moral Appraisability. Puzzles, Proposals and Perplexities*, Oxford University Press, Oxford/New York.

Han, Byung-Chul (2021), *No-cosas. Quiebras del mundo de hoy*, trad. esp. J. Chamorro Mielke, Taurus, Barcelona.

Huizinga, Johan (2010), *Homo ludens*, trad. esp., E. Ímaz, Alianza Editorial/Emecé, Madrid.

Ienca, Marcello, Roberto Andorno (2017), "A New Category of Human Rights: Neurorights". Disponible en: http://blogs.biomedcentral.com/bmcblog/2017/04/26/new-category-human-rights-neurorights/.

Illich, Ivan D. (1992) "L´alfabetizzazione informática e il sogno cibernético", en: Illich, I. D, (ed.), *Nello specchio del passato*, RED Edizioni, Milano.

Jakobs, Gunther, Miguel Polaino Orts (2009), *Derecho penal del enemigo: fundamentos, potencial de sentido y límites de vigencia*, Bosch, Barcelona.

Kapitan, Tomis (2000), *Autonomy and Manipulated Freedom, Philosophical Perspectives,* en: Action and Freedom, 14, 81-103.

Laporta San Miguel, Francisco (1987), "Sobre el concepto de derechos humanos", en: *Doxa* 4, 23-46.

Lettieri, Nicola (2020), *Antigone e gli algoritmi. Appunti per un approccio giusfilosofico*, Mucchi, Modena.

López Hernández, Hernán (2021), "Neuroderecho, neuroabogado, neurojusticia: una realidad innegable", en: Barona Vilar, S. (ed.) *Justicia algorítmica y neuroderecho. Una mirada interdisciplinar*, Valencia, Tirant lo Blanch, 87-108.

Marcus, Steven J (2002), *Neuroethics. Mapping the Field*, The Dana Press, New York.

Mele, Alfred R. (1995), *Autonomous Agents. From Self-Control to Autonomy*, Oxford University Press, Oxford/New York.

Moore, Schuyler (2021). "Law in the Metaverse", en: *Forbes*. Disponible en: https://www.forbes.com/sites/schuylermoore/2021/12/22/law-in-the-metaverse/?sh=2a431fab45d1

Morente Parra, Vanesa (2021), "La inteligencia híbrida: ¿hacia el reconocimiento y garantía de los neuroderechos? en: Llano Alonso, Fernando H, Joaquín Garrido Martín (ed.), *Inteligencia Artificial y Derecho. El jurista ante los retos de la era digital*, Cizur Menor (Navarra), Thomson Reuters Aranzadi, 259-277.

Ortega y Gasset, José (2006a). *Ensimismamiento y alteración* (1931), en: *Obras completas. Tomo V (1932/1940)*, Fundación José Ortega y Gasset/Taurus, Madrid, 529-550.

— (2006b), *Meditación de la técnica* (1931), en: *Obras completas. Tomo V (1932/1940)*, Fundación José Ortega y Gasset/Taurus, Madrid, 551-605.

Pérez Luño, Antonio-Enrique (1987), "Concepto y concepción de los derechos humanos: (Acotaciones a la ponencia de Francisco Laporta)", en: *Doxa*, 4, 47-66.

— (2006), *La tercera generación de derechos humanos*, Thomson/Aranzadi, Cizur Menor, Navarra.

— (2012), *Los derechos humanos en la sociedad tecnológica*, Editorial Universitaria, Madrid.

— (2018) (12ª ed.), *Derechos humanos, Estado de Derecho y Constitución* (1984), Tecnos, Madrid.

Pietropaoli, Stefano (2019), "Cyberspazio. Ultima frontera dell´inimicizia? Guerre, nemici e pirati nel tempo della rivoluzione digitale", en: *Rivista di Filosofia del diritto*, 2, 379-400.

Sartori, Giovanni (2018), *Homo videns. La sociedad teledirigida*, trad. esp., A. Díaz Soler, De Bolsillo, Barcelona.

Sententia, Wrye (2004), "Neuroethical Considerations: Cognitive Liberty and Converging Technologies for Improving Human Cognition", en: *Annals of the New York Academy of Science,* 1013, 221-228.

Sommaggio, Paolo, Marco Mazzocca, Alessio Gerola, Fulvio Ferro (2017), "Cognitive liberty. A first step towards a human neuro-rights declaration", en: *BioLaw Journal-Rivista di BioDiritto,* 3, 27-45.

Taylor, Charles (2020), *Fuentes del yo. La construcción de la identidad moderna*, trad. esp., A. Lizón, Paidós, Barcelona-Buenos Aires-México.

Taylor, J. Sherrod, J. Anderson Harp, Tyron Elliot (1991), "Neuropsychologists and neurolawyers", en: *Neuropsychology*, 5 (4), 293-305.

Vantin, Serena (2021), *Il diritto antidiscriminatorio nell´era digitale. Potenzialità e rischi per le persone, la pubblica Amministrazione, le imprese*, Wolster Kluwer, Milano.

Vašák, Karel. (1979), *Pour les droits de l´homme de la troisième génération*, Strasbourg, Institut International des Droits de l´Homme.

— (1990), "Les différents catégories des droits de l´homme", en: A. Lapeyre, F. De Tinguy y K. Vašák (eds.), *Les dimensions universelles des Droits de l´Homme*, Unesco-Bruylant, Bruxelles, 1-11.

Warwick, Kevin (2012), *Artificial Intelligence: The Basics*, Routledge, London-New York.

— (2015), "The Disappearing Human-Machine Divide", en: *Beyond Artificial Intelligence. The Disappearing Human-Machine Divide*, J. Romportl, E. Zackova y J. Kelemen (eds.), Cham-Heidelberg-New York-Dordrecht-London, Springer, 1-11.

Yuste, Rafael, Sara Goering *et al.* (2017 November 9), "Four ethical priorities for neurotechnologies and AI", en: *Nature,* 551, 159-163.

— (2021), Yuste, Rafael, Jared Genser, Stephanie Herrmann, "It´s Time for Neurorights", en: *Horizons. Journal on International Relations and Sustainable Development*. "The (Not So) Roaring Twenties"? Issue,18, 154-164.

CAPÍTULO IX

EN PRIMERA PERSONA. UN RÉQUIEM POR EL DERECHO DE LA ERA DIGITAL

STEFANO PIETROPAOLI
Universidad de Florencia
stefano.pietropaoli@unifi.it

Oro supplex et acclinis,
Cor contritum quasi cinis,
Gere curam mei finis

1. INTROITUS

En las páginas siguientes, el derecho del que se habla es el "derecho humano", hecho por seres humanos y para seres humanos. Este derecho humano no coincide, por supuesto, con el llamado "derecho natural", pero tampoco se corresponde con lo que suele entenderse por la expresión "derecho positivo": es, sencillamente, "derecho artificial". El derecho es artificial en la medida en que es "representación": una representación teatral -siempre trágica y consciente de ser también una representación sagrada- representada por actores que se mueven en un escenario con ropas y máscaras.

En el drama del derecho no falta la fuerza visual: la máscara, el traje, el gesto del actor. Sin embargo, esto cede al poder del sonido: al poder de la palabra y del silencio. El artificio humano que marca el derecho es la palabra. Las palabras del derecho son hechos, o más bien artefactos, resultado de un arte, una técnica y una ciencia que son obra del intelecto humano. Los juristas trabajan con palabras: elaboran conceptos, inventan estrategias argumentativas, atribuyen incesantemente significados a los enunciados lingüísticos. Las palabras no son entidades naturales, no se pueden tocar, no tienen corporeidad. Sin embargo, tienen una "materialidad" propia. Una vez pronunciadas, se convierten en "hechos": hechos que cambian la realidad natural de quienes están sentados en la audiencia.

El derecho se ha convertido en una ciencia -un conocimiento autónomo- por su capacidad de desarrollar su propio aparato conceptual. El derecho vive en conceptos y definiciones. Y esto no es para encerrarse en la jaula dorada de un sistema formal abstracto, sino porque los conceptos y las definiciones son las lentes con las que mira el mundo. Lo que llamamos "asesinato" no existe fuera del escenario del derecho. Un hecho como el asesinato de un hombre por otro hombre puede ciertamente ocurrir, pero lo relevante en el plano jurídico es la calificación de ese hecho concreto, capaz de generar consecuencias que afecten

concretamente a la realidad natural. Y son consecuencias muy diferentes, dependiendo de si ese hecho concreto se "figura" como, precisamente, un asesinato, o como la ejecución de una sentencia, como una acción de guerra, etc. Esta calificación, cuyo resultado está destinado a afectar a la carne viva de alguien, tiene lugar en el escenario del derecho, en una dimensión artificial, distinta de la realidad natural.

Las máscaras escénicas de los actores del derecho siempre se han llamado personas. La época actual, surcada por proyectos cada vez más radicales de dataficación y computabilidad integral, nos obliga a cuestionar el significado de este concepto, y en particular qué relación existe hoy entre la máscara y el actor, entre la persona y el ser humano.

Se trata, una vez más, de una relación trágica. El destino del derecho depende de la respuesta a esta pregunta (Campione, 2020). Estamos en el umbral de la salida definitiva del nexo entre hombre y persona, de la superación de lo humano tal y como lo hemos entendido y malentendido, de la entrada en una era habitada por personas no humanas o no del todo humanas. Una vez cruzado ese umbral, es posible que ya no encontremos un escenario en el que las mil máscaras del derecho sigan siendo utilizadas por actores humanos. Entonces habremos entrado en la era de una forma de derecho que no es humana, o que ya no es humana, o simplemente una era sin derecho. Como enseña el Digesto: "hominum causa omne jus constitutum". Borra a los humanos por un momento, imagina una realidad natural sin seres humanos: ¿puedes seguir viendo el escenario y escuchando la voz del derecho?

2. SEQUENTIA

Como es sabido, en la lengua materna del derecho *persona* significa máscara (Canale, 2015). Según una reconstrucción no exenta de incertidumbres (me remito a la interpretación de Gavius Bassus, recordada por Gellius in *Noctes Atticae*, 5.7.2), el término *persona* indicaba originalmente en latín la máscara teatral que llevaban los actores para intensificar su voz, haciéndola per-sonar, es decir, resonar para que pudiera ser escuchada incluso por los espectadores más alejados del escenario (Bettini, 2000). Hay buenas razones (incluso filológicas) para dudar de la validez de esta reconstrucción. Ahora bien, sigue siendo cierto que persona indicaba la máscara del actor, la evocación de un personaje representado por un intérprete. Este actor está al mismo tiempo desvinculado y, sin embargo, inexorablemente ligado al personaje que interpreta. La máscara es la representación de una figura que se hace presente en su ausencia (Pizzorno, 2007).

El término *persona* se introduce en el ámbito jurídico en el mismo momento en que la ciencia jurídica romana da sus primeros pasos. Desde el

origen del *jus*, *persona* designa la máscara con la que un actor -un actor humano- entra en el escenario del derecho. "Personam habere" (pero también "gerere", "suscipere", "sustinere") no significa otra cosa que llevar una máscara con la que es posible representar un papel en el escenario jurídico. La persona se revela así inmediatamente como un artificio, como una construcción, un artefacto, una creación del intelecto humano y de la ciencia jurídica en particular.

Teóricamente, la *persona* debe mantenerse separada de los conceptos que también la atraviesan constantemente y que a veces se superponen a ella: la subjetividad, la capacidad de obrar y la capacidad jurídica, la humanidad, el individuo. El pensamiento jurídico romano puede ayudarnos a distinguir estos diferentes niveles (Stolfi, 2007). En Roma nació el derecho tal y como lo entendemos hoy en día, se construyó el escenario sobre el que actuaban las *personae*, surgió la conciencia de la artificialidad de este instrumento en todo su poder (Thomas, 1998). Pero en Roma no hay rastro de "sujetos de derecho" (y "derechos subjetivos") ni de "capacidad de obrar". Es necesario resistirse a la tentación de buscar en la experiencia jurídica romana lo que sencillamente todavía no podía estar ahí, en la búsqueda desesperada de respuestas a "cuestiones cargadas de modernidad" (Stolfi, 2005, 398).

En la partición entre *personae, res* y *actiones*, la centralidad de la persona en el derecho romano es afirmada de forma tajante por Gayo (Ga. 1.8: "Et prius videamus de personis"), retomada por muchos juristas durante el Principado y recogida por las *Institutiones* de Justiniano (I., 1.2.12: "Ac prius de personis videamus"). Pero esta necesidad de pasar de la persona también se refleja inmediatamente en la evocación de una figura que revela toda la ambigüedad de este artificio: el esclavo.

Según la célebre distinción de Gaius (Gai. 1.9), la persona es un atributo tanto de los libres como de los esclavos: "Et quidem summa divisio de iure personarum haec est, quod omnes homines aut liberi sunt aut servi". Los esclavos, por tanto, son hombres -y esto ya es algo sobre lo que reflexionar- y hombres que pueden ser representados por una máscara en el escenario del derecho. Pero esto no es suficiente: los esclavos no sólo pueden ser suplantados por los libres, sino que también pueden suplantar a los propios libres. El actor sui juris es una persona que se interpreta a sí misma. El esclavo es un actor *alii juris*, que puede llevar la máscara de la persona libre. Por lo tanto, la máscara del esclavo puede ser llevada por el amo, pero la máscara del amo también puede ser llevada por el esclavo, en caso de que el *dominus* desee autorizarlo a actuar en su nombre (para, por ejemplo, celebrar un contrato en su nombre). El esclavo romano, como se repite a menudo, es una "cosa". Pero también es una "persona". No hay contradicción, por más que esta afirmación sea hoy

inaceptable a los ojos de quienes piensan en la "persona humana" del constitucionalismo del siglo XX.

Actores que intercambian sus papeles, actores que interpretan varios roles, actores que hablan o permanecen en silencio: todo esto sucede en el escenario del derecho. Y es precisamente la posibilidad que tiene un mismo ser humano de encarnar varios roles en un mismo escenario lo que explica la apropiación del concepto de persona por parte de la teología: basta pensar en las interminables discusiones que atravesaron los primeros siglos del cristianismo y que condujeron a la elaboración del dogma trinitario, según el cual la sustancia (única) de Dios opera a través de tres personas distintas (Milano, 1984).

La teología cristiana -más claramente con Tomás, pero hay muchos autores que deben ser recordados- inició un proceso de asimilación del concepto de "persona", que pronto empezó a designar cualquier entidad racional (divina o humana). Desde esta fractura, inseparablemente unida a una concepción de la naturaleza como orden racional creado por Dios (por cierto: una "naturaleza" muy diferente a la de los romanos), llegaremos a la identificación entre persona y ser humano: y en esta perspectiva se moverá no sólo el personalismo cristiano del siglo XX, sino también el gran número de teorías "seculares" que se han apoyado en el potencial expresado por la dignidad de la persona humana. Pero, repitámoslo: en su acepción puramente jurídica, la persona no coincide con el ser humano, hasta el punto de que la persona puede "nacer" antes que él (como el *infans conceptus*), al igual que puede "morir" después que él (piénsese en la herencia yacente o en la muerte presunta; y, en este último caso, también puede "morir" antes que él).

También hay que registrar otro cambio decisivo en cuanto a la relación entre las personas y las cosas. Mientras que en el derecho romano una figura como la del esclavo -capaz de ser tanto persona como res- era absolutamente justificable, en el derecho medieval ya no hay lugar para figuras intermedias entre las personas y las cosas. Obviamente, esta cuestión está relacionada con la distinción cristiana entre alma y cuerpo. En el derecho romano (que conocía las *res sacrae* pero no las *personae sacrae*) detrás de la persona había un ser humano con cuerpo y alma, pero la persona no participaba de ninguna sacralidad. En el derecho medieval, la persona se sacraliza porque se fusiona y confunde con un individuo racional provisto de alma.

Precisamente a partir de estas desviaciones, el intercambio entre la ciencia jurídica y la teología en torno al concepto de persona estaba destinado a perpetuarse durante toda la Edad Media. Pensemos en el tema de la Iglesia como *corpus mysticum* y en la reflexión paralela sobre el concepto de *fictio juris* (Thomas, 1995, 21), elaborado en particular por Bartolo da Sassoferrato, Paolo

di Castro y Giovanni da Imola, y aún más al tema de la representación/representación (Hofmann, 1974), que implica directamente la cuestión de la *persona ecclesiae*; por ejemplo, el sacerdote, al celebrar el misterio eucarístico, pronuncia una fórmula en la que el sacrificio se ofrece "non tantum in sua, sed in totius Ecclesiae persona" (el Papa Innocentio III se ejercitó en el tema: cf. *De sacro altaris mysterio,* en *Patrologia cursus completus*, serie Latina, accurante I.P. Migne, París 1844 y siguientes, CCXVII, col. 844 C, lib. III, cap. 5).

Es precisamente con los juristas medievales (Bartolo y Baldo en particular) que la persona ya no se limita a "llevarse" sino que desempeña explícitamente una función de "representación".

El término "persona" comienza a tener un alcance semántico diferente: ya no indica (sólo) estar en el lugar de otro en el escenario del derecho, sino que ahora indica la unidad ficticia de representante y representado. La historia de lo que antes se llamaba *persona moralis* y lo que ahora llamamos "persona jurídica" es atribuible a este cambio: la persona jurídica es una unidad de seres humanos y bienes, o incluso sólo una masa de bienes, que puede estar en el escenario del derecho gracias al "representante" (pensemos en la *universitas*). Y es en esta perspectiva inédita que los juristas parecen cada vez menos conscientes de que incluso la "persona física" es un artificio, tan incorpóreo como la "persona jurídica" (Ranieri, 2020).

En este surco se encontrarán una serie de reflexiones que, habiendo atravesado la Baja Edad Media, desembocarán en la primera edad moderna con una concepción naturalista según la cual el cuerpo político, formado por una multitud de sujetos, encontrará la unidad en la única persona del soberano. El Estado, nuevo protagonista de la modernidad, es una "persona" como centro de imputación de los intereses de una entidad colectiva. Es persona y, por tanto, artificial: Oswald Hilliger, comentando un pasaje de Donellus, puede afirmar que "hombre" es una palabra de naturaleza, "persona" es una palabra de derecho ("homo naturae, persona juris civilis vocabulum": *Hugonis Donelli ... Opera omnia. Commentariorum De iure civili tomus primus - [duodecimus] cum notis Osualdi Hilligeri. Accedunt summaria, & castigationes theologicae*, Lucae 1762-1770).

Pero en este punto hay que aclarar un posible malentendido. El mayor teórico del Estado, Thomas Hobbes, como es bien sabido (De homine, XI, I, y Leviatán, I, XVI y XXVI), distingue entre "personas naturales", que hablan y actúan "en primera persona", y "personas artificiales", que hablan y actúan en nombre de otra persona. Por ejemplo, el Estado en esta perspectiva se presenta como *persona civilis*, que "actúa" los sujetos (Amendola, 1998).

El problema que plantea la elección terminológica de Hobbes, evocada por una expresión como "persona natural", es evidente: la persona es artificial,

pero "natural" al mismo tiempo. Sin embargo, no se trata de un oxímoron: cuando Hobbes utiliza el adjetivo "natural" en referencia a la persona, no quiere referirse a un estatus innato perteneciente al ser humano como tal. Más bien pretende aludir a la posible -no necesaria- identidad entre el portador de la máscara y la misma máscara. En otras palabras, una persona "natural" es el actor que se interpreta a sí mismo, sin ninguna referencia a un estatus presocial.

En el estado de naturaleza no hay personas, sólo hombres. Los hombres se ponen la máscara de persona sólo después de entrar en el estado civil. Antes de la creación del Estado, no puede haber persona porque no hay derecho: se confirma así el vínculo indisoluble (y artificial) entre persona y derecho. Por lo tanto, la personalidad jurídicamente relevante no es un atributo de todos los seres humanos como miembros de la especie humana. Es el Estado el que confiere este estatus a sus súbditos. Y siempre es el Estado el que puede ampliar o restringir este reconocimiento independientemente de la pertenencia biológica a la especie humana.

En el curso de la modernidad, la parábola del "sujeto de derecho" se impone inexorablemente, y paralelamente se desarrolla el discurso de los "derechos subjetivos". El concepto de persona se emplea cada vez más para designar esta subjetividad. Y es en este mecanismo de referencias recíprocas donde la persona se convierte en un dispositivo de subjetivación y, al mismo tiempo, de objetivación y sujeción (Bazzicalupo, 2013). En otras palabras, el concepto de persona -originalmente marcado por la separación entre la máscara y el rostro- se presta ahora a definir no sólo umbrales de división entre la subjetividad jurídica y el cuerpo, sino también formas de separación entre seres vivos plenamente humanos y seres bestiales.

Así, desde finales del siglo XVIII, el dispositivo jurídico de la persona se utiliza para superar la abstracción formal del sujeto de derecho, incluyendo su dimensión corpórea (Baud, 1993). Pero es precisamente este reconocimiento de la corporeidad del sujeto lo que ha permitido imaginar formas de "vida cualificada", en las que la persona es el componente capaz de dominar la parte animal. La posibilidad de distinguir a los seres vivientes en jerarquías que reflejen esta diferente capacidad de control corresponde a la identificación de sujetos plenamente humanos (personas) y sujetos que se alejan progresivamente de este estatus, hasta llegar a lo no humano. Además, como han argumentado Roberto Esposito (Esposito, 2007) y Alberto Moreiras, el ser-persona de unos siempre ha supuesto el no-ser-persona de otros: la no-persona es siempre "el reverso de la persona, su exigencia oculta". No hay persona sin no-persona, como no hay máscara sin rostro de carne y hueso, ni apariencia sin presencia fundante, ni ficción sin realidad, al menos imaginaria, de fondo" (Moreiras, 2008).

Comenzó así a ahondar en el subsuelo de lo que -con razón y paradójicamente al mismo tiempo- se llamará la "era de los derechos" (Bobbio, 1997), una criatura monstruosa cuya perversión emergerá en toda su violencia con el exterminio judío. A finales del siglo XIX, la ciencia jurídica ya había iniciado un proceso, sigiloso y a menudo no del todo consciente, que la llevó primero a distinguir entre personas en sentido pleno y personas con capacidad jurídica limitada, y luego a negar la condición de persona a los no arios (Rottleuthner, 1983). La Shoah se erige, así como la culminación de un largo viaje en el que el conocimiento biológico, lentamente absorbido por la teoría jurídica, se convirtió en la base de las prácticas de la "tanatopolítica" nazi. Las justificaciones basadas en la biología y la genética pueden ahora utilizarse para superar la abstracción de la persona, derribando su fisonomía como sujeto racional y centro de imputación de derechos y relaciones jurídicas, con el resultado de reducir al hombre a su sustancia biológica: una sustancia estriada, que hace posible las jerarquizaciones basadas en la raza.

Después de la catástrofe, hubo una fuerte necesidad de devolver la centralidad al ser humano como individuo racional, dotado de derechos inalienables y caracterizado por una dignidad que no puede serle arrebatada de ninguna manera. Y, como ya se ha dicho, el término "persona" fue objeto de una recuperación, llevada a cabo simultáneamente en la estela de las distintas culturas que, de diversas maneras, ya habían hecho uso de él: ningún término se consideró más adecuado para apoyar las nuevas luchas jurídicas, políticas, económicas y sociales (pienso en particular en autores del ámbito católico como Maritain y Mournier).

La consagración de este renacimiento puede considerarse la *Declaración Universal de los Derechos Humanos* de 1948, que se refiere a la dignidad de la "persona humana" como titular de derechos inalienables. Si bien este renacimiento de la persona ha producido sin duda resultados positivos (Rodotà, 2006), también es cierto que en sí mismo la presunta universalización de la persona y de la humanidad obedece a lógicas políticas que generan fatalmente nuevas formas de discriminación, como revela la retórica de las "guerras humanitarias" que marcaron la última parte del siglo pasado (Zolo, 2006).

Como ha argumentado Roberto Esposito, el concepto de persona está marcado por una paradoja: parece destinado a reproponer eternamente la polaridad entre cuerpo y alma, y a producir así una brecha entre animalidad y racionalidad que puede conducir potencialmente a una gradualización de la pertenencia a la humanidad. En otras palabras, el concepto de persona parece ser incapaz de ofrecer nada más que aquello para lo que fue desarrollado originalmente. Era y sigue siendo un producto artificial del intelecto humano,

que remite a una distinción necesaria entre la máscara y el actor que la lleva. En esta perspectiva, hay que subrayar que la ciencia jurídica misma ha entendido sistemáticamente que las personas -que son siempre "jurídicas" en el sentido de tener un estatuto jurídicamente relevante- son entidades que no coinciden necesariamente con los seres vivientes individualmente identificados: las asociaciones, las sociedades, las fundaciones, son sólo algunos ejemplos de entidades colectivas a las que el derecho reconoce un estatuto personal.

Hans Kelsen sostenía que, en el plano teórico, la persona -"física" o "jurídica", según el léxico jurídico tradicional- no "tiene" derechos y deberes, sino que, más propiamente, "es" este complejo de deberes jurídicos y derechos subjetivos, cuya unidad se expresa figurativamente en el concepto de persona. La persona no es más que la personificación de esta unidad (Kelsen, 1960). Por lo tanto, no hay ningún escándalo jurídico en reconocer la personalidad de sujetos que ni siquiera coinciden con las entidades colectivas humanas: esta máscara puede utilizarse para representar legalmente a animales distintos del hombre, pero también a criaturas no animales, e incluso a autómatas dotados de inteligencia artificial. Sin embargo, hay que ser conscientes de que, en un movimiento absolutamente especular y sin embargo coherente, el derecho ha excluido y excluye el reconocimiento de la personalidad a los seres vivos biológicamente pertenecientes al género humano. La "persona física" surge en el momento del nacimiento del ser humano. El niño no nacido no tiene máscara. Y, como él, el casi-muerto, el esclavo, la mujer, el niño, el loco, el judío, el indígena: son sólo algunos ejemplos de una historia también hecha de exclusiones y discriminaciones, en la que se han diferenciado y se diferencian las personas, las cuasi-personas y las no-personas.

El dispositivo de la persona no puede escapar a esta lógica. Y, entonces, puede ser útil intentar esbozar una taxonomía de la persona. En primer lugar, podemos identificar a las personas "naturales" (atención: en el sentido ya mencionado de máscaras artificiales que representan entidades que existen concretamente en la realidad natural). Entre ellos se encuentran, por supuesto, los seres pertenecientes a la especie *homo sapiens sapiens*, ya sea como individuos o como grupos de individuos. Pero otras criaturas orgánicas, como los animales que no son humanos, también pueden incluirse sin ninguna duda. Ni siquiera tengo la oportunidad de mencionar aquí el impresionante debate sobre los derechos de los animales de las últimas décadas. El hecho es que, hoy en día, sin ningún tipo de escándalo, la máscara de los animales se usa en las salas de justicia de la mayoría de los países del mundo.

Esta misma posibilidad se reconoce cada vez más para otras entidades orgánicas, como las plantas. Entre los muchos casos, es famoso el del roble blanco del coronel William Henry Jackson, que en 1832 cedió a su amada planta

la plena propiedad de sí misma y del terreno situado a dos metros de su tronco[1]. A la máscara arbórea se han unido en los últimos años las de otras entidades que (con una aproximación que espero se me perdone) en su unidad diría que son "inorgánicas", aunque rebosantes de vida. El 15 de marzo de 2017, el Parlamento neozelandés concedió el estatus de "persona" al río Whanganui, confiando su representación a la comunidad maorí Iwi. Poco después, el Tribunal Superior del estado indio de Uttarakhand concedió personalidad jurídica primero al Ganges y luego a los glaciares del Himalaya.

Así, en el escenario del derecho aparecen máscaras de hombres, animales, árboles, ríos y montañas. Pero eso no es suficiente. Además de todas estas personas "naturales", encontramos personas "sobrenaturales". En 2010, de nuevo en la India, el Tribunal Superior de Allahabad autorizó la comparecencia de Bhagwan Sri Ram Virajman, una deidad local, para que se opusiera, a través de su representante, a la construcción de un edificio en lo que tradicionalmente se consideraba su *janmabhoomi* (lugar de nacimiento). El 19 de noviembre de 2019, comentando la conclusión del largo proceso judicial, el New Dehli Business Standard tituló triunfalmente: «Ayodhya verdict: Bhagwan Sri Ram Lalla Virajman emerges the clear winner. Rejecting challenges to the lawsuit of the deity, the SC said the possession of "inner and outer courtyards" be handed over to the trust or to the body so constituted».

Por otra parte, no es necesario molestar a las deidades exóticas para identificar casos de atribución de personalidad jurídicamente relevante a entidades "sobrenaturales", que también están bien presentes en el corazón del pensamiento occidental. La figura hobbesiana del Leviatán -que es simultáneamente dios, hombre, animal y máquina- es posiblemente la que mejor revela el poder de una persona "artificial" que está más allá de la dimensión física y que también se desplaza de la suma de los cuerpos de sus súbditos para convertirse en un mecanismo animado y en una divinidad al mismo tiempo, si bien divinidad mortal (Schmitt, 1938).

Sin embargo, en todos estos casos, la máscara la llevan actores humanos. Son seres humanos que se hacen pasar ahora por un animal, ahora por una montaña, ahora por una deidad. Y así nos enfrentamos al verdadero problema que nos plantean las tecnologías digitales. No se trata en absoluto del reconocimiento de formas de personalidad jurídicamente relevantes a los autómatas, robots, máquinas dotadas de inteligencia artificial. Como creo que ya está claro, se trata de un falso problema: el derecho ha inventado la persona para representar todo aquello que es representable y, por tanto, no hay ningún

[1] Cf. *This Tree Owns Itself*, in *Athens Daily Banner*, June 16, 1901, 8 (disponible en versión digital en https://gahistoricnewspapers-files.galileo.usg.edu/lccn/sn89053947/1901-06-16/ed-1/seq-8.pdf).

problema jurídico (sino sólo político) en reconocer la personalidad a un androide social, como, por ejemplo, 'Sophia', desarrollada por Hanson Robotics Limited, a la que se le concedió la ciudadanía saudí el 25 de octubre de 2017 y que el 21 de noviembre del mismo año se convirtió en la primera entidad artificial en recibir un título de Naciones Unidas a través de su nombramiento como Campeón de la Innovación del Programa de Desarrollo de la ONU. El problema es que, por primera vez en la historia, la máscara con la que se entra en la escena de la ley no la pueden llevar los seres humanos.

Si, repito, no hay ningún problema jurídico en atribuir personalidad a un androide y, por tanto, en que un actor humano actúe en el escenario del derecho bajo la apariencia de un androide, cuestión distinta se plantea cuando el androide toma la palabra "en primera persona". El androide, una entidad artificial, se convierte así, paradójicamente, en una "persona natural" en el sentido atribuido a la expresión por Hobbes. La pregunta es: ¿Puede el no humano actuar en el escenario del derecho? ¿Puede el derecho ser "interpretado" por actores no humanos? ¿Seguiría siendo derecho?

3. LACRIMOSA

El derecho, como hemos visto, puede establecer "personas" de cualquier tipo. Si, desde esta perspectiva, el problema del reconocimiento de la personalidad jurídica de un androide dotado de I.A. es una cuestión fácilmente superable, el problema de una entidad artificial que no está representada por un ser humano, pero que pretende actuar jurídicamente en primera persona, es de una gravedad completamente distinta.

Podríamos, tal vez, tener la tentación de negar esta posibilidad, y atrincherarnos en el reducto de el "derecho humano ", creado por los hombres para los hombres, con mil máscaras posibles, pero todas llevadas por seres humanos. Sin embargo, aunque esta respuesta pueda parecer salvadora, el problema persiste. Permítanme al menos mencionar dos umbrales críticos para la ciencia jurídica.

La primera consideración: en esta época de revolución digital, la personalidad humana tiene cada vez más que ver con los "datos". No quiero decir que ahora esté totalmente desvinculado del cuerpo. Pero me parece claro que estamos asistiendo a los primeros pasos de un proceso de hibridación entre el cuerpo físico (en la visión clásica del habeas corpus) y la proyección digital de las actividades intelectuales y las prácticas sociales que se refieren a ese cuerpo. En este sentido hoy se habla de *habeas data* (Pietropaoli, 2020). Muchas constituciones contemporáneas se refieren a los derechos inviolables del hombre, tanto como individuo como en las formaciones sociales donde se desarrolla su personalidad. Este es uno de los temas centrales en la actualidad:

el lugar de las formaciones sociales donde se desarrolla nuestra personalidad está representado cada vez más por el llamado ciberespacio. En este extraño espacio que no tiene ubicación física, se juega por tanto nuestra personalidad, que obviamente no puede ser otra cosa que una personalidad traducida en datos. Estamos asistiendo a una dataficación de la persona. Personalmente, no creo que la digitalización en sí misma sea un problema. En cambio, creo que la idea de reducir la persona humana en su totalidad a datos, a elementos computables, es un problema.

Y aquí llegamos a la segunda consideración, que evoca un problema aún más radical. Las tecnologías que nos desafían hoy en día, y que parecen socavar el ya frágil concepto de "humano" y la relación entre el hombre y la persona, son las de la llamada mejora humana (Palazzani, 2015). El problema antropológico que subyace a los desarrollos de estas tecnologías se refiere a la reflexión sobre los límites de la manipulación del hombre y de la humanidad ante posibles intervenciones en la progresiva artificialización de lo humano o antropomorfización de la tecnología. Son escenarios que dibujan horizontes en los que lo artificial se asemeja cada vez más a lo natural, tendiendo a fundirse y "mezclarse" con él, de manera que la diferencia entre hombre y máquina se anula, en una simbiosis entre hombre y tecnología, entre vida orgánica e inorgánica, entre vida animada e inanimada. Como ha escrito magistralmente Remo Bodei, parafraseando el Evangelio de Juan, el Verbo se hizo máquina, el espíritu también sopla en lo inorgánico, y la razón y el lenguaje, objetivados en forma de algoritmo, habitan en cuerpos no humanos, creando una 'humanidad aumentada' (Bodei, 2018). Esta perspectiva alude explícitamente a la transformación y "mejora" del hombre. No se trata de una función correctiva para "reparar" el cuerpo (para corregir los resultados de una enfermedad o lesión, por ejemplo), sino de la potenciación de las capacidades fisiológicas -físicas y mentales- del ser humano. Los miembros, los órganos, la mente forman parte de un mecanismo, de una máquina que puede ser manipulada y así perfeccionada.

El mundo del hombre, hecho de carne y hueso, se está fusionando con el mundo de la máquina, hecho de bits y silicio. Se trata de un movimiento doble y convergente. Por un lado, el hombre utiliza cada vez más prótesis artificiales de todo tipo, que injertadas en su cuerpo le permiten desarrollar capacidades totalmente nuevas. Por otra parte, las máquinas están adquiriendo capacidades y cualidades humanas: una racionalidad y autonomía que imitan a las humanas (si no una verdadera "inteligencia"), pero también una fisicalidad diferente, basada en la posibilidad de aprovechar los tejidos orgánicos. Y si la creación de ciborgs evoca una visión distópica todavía aparentemente lejana, la creación de "ordenadores moleculares" se está convirtiendo en una realidad.

Lo orgánico y lo sintético se fusionan, en una nueva simbiosis que no puede dejar indiferente a nadie. El hombre se mecaniza, la máquina se humaniza. La baraja de los límites establecidos por la naturaleza humana se está volviendo a barajar. Lo paradójico es que este horizonte fue diseñado para obviar el azar de la vida, marcado por la casualidad y el destino, pero al abrir estas nuevas perspectivas ya no permite ninguna predicción fiable.

La vieja humanidad a la que podíamos apelar refiriéndonos a una dimensión biológica se ha quedado "obsoleta" (Anders, 2011), y da paso a los escenarios de lo transhumano y lo posthumano (Llano Alonso, 2018). No hay oportunidad de abordar aquí el significado de estas expresiones. Pero lo que nos basta con señalar en estas páginas es que lo humano muestra ahora límites cambiantes bajo la presión de la oportunidad de fabricar una nueva especie, bajo la bandera de la idea de que la mejora humana no es más que una fase de la evolución en la que la tecnología permite sustituir la selección natural por una selección deliberada por el hombre mismo. Es en este marco donde encontramos las corrientes tecnofílicas más extremas que sueñan un futuro de liberación de todas las limitaciones biológicas que marcan la condición humana, con vistas a una nueva condición tecnohumana de emancipación del hombre de su propia naturaleza en la evolución de un nuevo ser "ya no" hombre, sino "otro" que el hombre y "más allá" del hombre. Según algunos autores, hemos llegado a una etapa en la que la evolución darwiniana está a punto de dar paso a una dinámica en la que el hombre toma las riendas de la evolución en sus propias manos y transmite directamente a su descendencia las modificaciones que considera oportunas. El perfeccionamiento representa, en este sentido, una fase del evolucionismo: la selección natural va a ser sustituida por la "elección deliberada" del proceso de selección, que permite alcanzar el mismo resultado más rápidamente. Aunque todavía no se conocen los posibles resultados negativos, bloquear ahora los avances en esta dirección significaría obstaculizar o impedir la posibilidad de acelerar la evolución de la humanidad. La mejora, desde esta perspectiva, se considera una "apuesta razonable" que debe llevarse a cabo. Es la teorización de la autoevolución y la mejora evolutiva (*enhancement evolution*) la que acorta los millones de años de progreso evolutivo. Las intervenciones de mejora sustituyen artificialmente a la selección natural y mejoran las condiciones físicas y sociales de los seres humanos, ejerciendo un control sobre el desarrollo futuro de la humanidad y contribuyendo al cambio radical de la naturaleza humana.

La aplicación no terapéutica de tecnologías específicas a la biología humana implica, pues, algo más que el mero restablecimiento del estado de salud o la funcionalidad "normal" del sujeto. En esta dirección se inscriben las tecnologías designadas por el acrónimo NBIC: nano (operaciones a escala atómica o molecular), bio (tecnologías aplicadas a los organismos vivientes),

info (técnicas que procesan y utilizan datos), cono (disciplinas que estudian el pensamiento en los campos de la neurociencia, la psicología y la lingüística). Estas tecnologías exigen un nuevo campo de reflexión que exige cuestionar los límites entre salud y enfermedad, lo normal y lo patológico, los fines de la medicina, el sentido de la curación, pero también el sentido de la naturaleza, la identidad humana y la justicia social en el contexto de la relación entre la tecnología y el ser humano y en la perspectiva de un "deber de responsabilidad" que la generación actual tiene hacia las generaciones futuras, pues la cuestión que se plantea ya no es si es moralmente aceptable interferir en los procesos naturales, sino si es legítimo utilizar las nuevas tecnologías que nos permiten remodelar las características de las generaciones futuras, de los animales no humanos y del medio ambiente.

Y si las tecnologías disponibles hoy en día parecen hacer por primera vez posibles esas intervenciones de forma concreta, la idea, el proyecto que hay detrás de su realización acompaña a toda la modernidad. A principios del siglo XVII, en el apéndice de la *Nueva Atlántida*, Francis Bacon ya imaginaba: "Prolongar la vida; retrasar la vejez; curar las enfermedades consideradas incurables; calmar el dolor; transformar el temperamento, la altura, las características físicas; fortalecer y exaltar habilidades intelectuales; transformar un cuerpo en otro; fabricar nuevas especies; hacer trasplantes de una especie a otra; crear nuevos alimentos utilizando sustancias que no se usan en la actualidad".

Así, por un lado, estas técnicas pueden permitir que alguien que ha sufrido una lesión cerebroespinal vuelva a caminar. Pero, por otro lado, pueden permitir que una persona normalmente capacitada a la que se le han implantado prótesis biónicas corra más rápido que un corredor olímpico de cien metros. Las operaciones de este segundo tipo han sido prefiguradas (y probablemente experimentadas) en el ámbito militar: la figura de un supersoldado, físicamente fuerte, con una vista y un oído excepcionales, sin emociones (y por lo tanto sin miedo), incapaz de sentir dolor y fatiga, es obviamente el sueño de cualquier general.

Esta mejora puede referirse no sólo a la dimensión estrictamente biológica (como la mejora física) sino también a la neurocognitiva, lo que se traduce en una mejora del rendimiento emocional y mental. Por lo tanto, no se trata sólo de potenciar a los atletas que puedan vencer a la competencia en una competición deportiva. Pero también estamos hablando de candidatos a una oposición para ser magistrados que, gracias a implantes cerebrales, pueden cargar durante la prueba todos los datos de la legislación, la jurisprudencia y la doctrina relevantes para responder a una determinada pregunta (Pietropaoli, 2018).

La cuestión que debe agitarnos es si todo lo que es tecnológicamente posible puede considerarse éticamente admisible, socialmente aceptable y legalmente permisible. De hecho, la exaltación de la tecnología corre el riesgo de hacer coincidir la mejora humana con la anulación de la voluntad subjetiva sobre la naturaleza objetiva. De ahí la idea de algunos estudiosos de contraponer la dimensión de la mejora a la del logro, es decir, la perspectiva de "adquisición, realización, logro, como desarrollo y realización del potencial naturalmente inscrito en el llegar a ser lo que se es, a través de un esfuerzo activo y un compromiso personal que permiten modificar las capacidades naturales de uno mismo mejorándose" (Palazzani, 2015, p. 41). En esta visión, lo que ofrece la tecnología no es la única forma de mejorar la humanidad: "empoderarse" implica, en cambio, el compromiso, el esfuerzo, la tensión individual, con el objetivo no de mejorar ciertas funciones humanas, sino de lograr un "crecimiento humano" general.

El problema, por tanto, ya no es sólo el de comprender la relación entre el hombre y la persona, y cuestionar la posibilidad de que la máscara de la persona pueda ser llevada por una criatura completamente artificial (Saporiti/Salardi, 2020): el problema aún más acuciante es entender si podría ser un sujeto híbrido el que llevara esa máscara, dotado de la capacidad de conocimiento protésico, es decir, de acceder a información que el individuo no conoce per se, sino sólo a través de implantes neurotecnológicos.

Si un escenario poblado de neuroprótesis que permitirán el uso de la tecnología informática y neurocientífica como ayuda permanente es cualquier cosa menos ciencia ficción, hay que cuestionar el nuevo perímetro que definirá el propio concepto de "humanidad" y el espacio de libertad de la "persona humana" como sujeto protegido. En el mundo de la humanidad mejorada, las tecnologías de mejora humana podrían delinear nuevos estatus legales de la personalidad, accesibles sólo a quienes tengan los recursos económicos para afrontar los costes. Así, habría una división (y discriminación) entre humanos mejorados y humanos despotenciados.

Por tanto, es necesario preguntarse, junto con Stefano Rodotà: ¿qué ocurre, sin embargo, cuando la innovación científica y tecnológica permite mejorar el rendimiento físico e intelectual? Si estas nuevas oportunidades se ofrecen de forma selectiva, si el acceso depende de los recursos financieros, llegamos a una sociedad de castas; se produce una reducción de la ciudadanía, que se convierte en censura; más dramáticamente, llegamos a una división humana, a un mundo que acepta la construcción de personas estructuralmente diferentes. ¿Debemos concluir que "el hombre está anticuado", como sugirió Günther Anders? ¿O debemos más bien retomar el hilo de la asociación entre

dignidad e igualdad, la única que puede evitar la separación radical entre las personas, la guerra entre humanos y posthumanos con cualidades diferentes?

Al intentar responder a esta pregunta, el filósofo del derecho no puede dejar de plantear otra pregunta: ¿a quién le habla el derecho? y, mejor aún, ¿quién dice el derecho? no se trata de qué, sino de quién. En el ámbito del derecho, el individuo humano no es una persona como tal, sino que pasa a serlo, como cualquier entidad colectiva, cuando el ordenamiento jurídico lo convierte en centro de imputación de derechos y relaciones jurídicas. Y aquí llegamos al punto. ¿A quién habla el derecho? ¿Y quién dice el derecho? Las personas. Es decir, máscaras. Máscaras escénicas para el escenario del derecho. Así que, en conclusión, la cuestión decisiva es si en la era de la inteligencia artificial esta máscara la seguirán llevando única y exclusivamente los hombres biológicamente pertenecientes a la estirpe humana.

Hubo un tiempo en el que habría sido relativamente sencillo dar una definición legal de "hombre" basada en esta identidad biológica. Pero justo cuando la biología entró en escena, fuimos testigos de las peores discriminaciones de nuestra historia. Ese tiempo ya ha pasado. Tenemos que entender ahora si el derecho puede hacer frente a un escenario sin precedentes: un escenario en el que la máscara de Sophia ya no será llevada por un ser humano, sino por una entidad no biológica sui juris; y en el que el actor que antes podíamos llamar humano, dejará de serlo en absoluto. ¿Llegará el día en que el jurista, con el corazón casi reducido a cenizas, pida a una entidad con inteligencia artificial que se encargue de su desaparición y la de su ciencia? *Lacrimosa dies illa*.

4. BIBLIOGRAFÍA

Amendola, Adalgiso (1998), *Il sovrano e la maschera. Il concetto di persona in Thomas Hobbes,* ESI, Napoli.

Anders, Günther (2011), *La Obsolescencia del hombre,* Pre-Textos, Valencia.

Baud, Jean-Pierre (1993), *L'affaire de la main volée. Une histoire juridique du corps,* De Seuil, Paris.

Bazzicalupo, Laura (2013), *Dispositivi e soggettivazioni,* Mimesis, Milano - Udine.

Bettini, Maurizio (2000), *Le orecchie di Hermes. Studi di antropologia e letterature classiche,* Einaudi, Torino.

Bobbio, Norberto (1997), *L'età dei diritti,* Einaudi, Torino.

Bodei, Remo (2018), *Dominio e sottomissione. Schiavi, animali, macchine, intelligenza artificiale,* Il Mulino, Bologna.

Campione, Roger (2020), *La plausibilidad del derecho en la era de la inteligencia artificial. Filosofía carbónica y filosofía silícica del derecho,* Dykinson, Madrid.

Canale, Damiano (2015), "Persona", en: Mario Ricciardi, Andrea Rossetti y Vito Velluzzi, *Filosofia del diritto. Norme, concetti, argomenti,* Carocci, Roma, 15-35.

Esposito, Roberto (2007), *Terza persona. Politica della vita e filosofia dell'impersonale,* Einaudi, Torino.

Hofmann, Hasso (1974), *Repräsentation. Studien zur Wort-und Begriffsgeschichte von der Antike bis ins 19. Jahrhundert,* Duncker und Humblot, Berlin.

Kelsen, Hans (1960), *Reine Rechtslehre,* Franz Deuticke, Wien.

Llano Alonso, Fernando H. (2018), *Homo Excelsior. Los Límites ético-jurídicos del transhumanismo,* Tirant lo Blanch, Valencia.

Milano, Andrea (1984), *Persona in teologia. Alle origini del significato di persona nel cristianesimo antico,* Edizioni Dehoniane, Bologna.

Moreiras, Alberto (2008), "La vertigine della vita", en: Bazzicalupo, Laura, *Impersonale. In dialogo con Roberto Esposito*, Mimesis, Milano-Udine.

Palazzani, Laura (2015), *Il potenziamento umano. Tecnoscienza, etica e diritto,* Giappichelli, Torino.

Pietropaoli, Stefano (2018), "Fine del diritto? L'intelligenza artificiale e il futuro del giurista", en: Dorigo, Stefano, *Il ragionamento giuridico nell'era dell'Intelligenza Artificiale*, Pacini, Pisa, 107-118.

— (2020), "Habeas data. I diritti umani alla prova dei big data", en: Sebastiano Faro, Tommaso Edoardo Frosini y Ginevra Peruginelli, *Dati e algoritmi. Diritto e diritti nella società globale*, Il Mulino, Bologna, 97-111.

Pizzorno, Alessandro (2007), "Saggio sulla maschera", en: Pizzorno, Alessandro, *Il velo della diversità. Studi su razionalità e riconoscimento,* Feltrinelli, Milano.

Ranieri, Filippo (2020), *L'invenzione della persona giuridica. Un capitolo nella storia del diritto dell'Europa continentale,* Giuffrè, Milano.

Rodotà, Stefano (2006), *La vita e le regole. Tra diritto e non diritto,* Feltrinelli, Milano.

Rottleuthner, Hubert (1983), *Recht, Rechtsphilosophie und Nationalsozialismus*, Steiner, Wiesbaden.

Saporiti, Michele y Salardi, Silvia (2020), "Perché l'IA non deve diventare Persona. Una critica all'ineluttabile 'Divenire antropomorfico' delle Macchine", en: Silvia Salardi y Michele Saporiti, *Le tecnologie 'morali' emergenti e le sfide etico-giuridiche delle nuove soggettività,* Torino, Giappichelli, 52-74.

Schmitt, Carl (1938), *Der Leviathan in der Staatslehre des Thomas Hobbes. Sinn und Fehlschlag eines politischen Symbols,* Hanseatische Verlagsanstalt, Hamburg.

Stolfi, Emanuele (2005), "I «diritti» a Roma", en: *Filosofia politica,* 3, 383-398.

— (2007), La nozione di «persona» nell'esperienza giuridica romana, en: *Filosofia Politica,* 21, 379-392.

Thomas, Yan (1998), Le sujet de droit, la personne et la nature. Sur la critique contemporaine du sujet de droit, en: *Le débat,* 100, 3, 85-107.

— (1995), Fictio legis. L'empire de la fiction romaine et ses limites médiévales, en: *Droits. Revue française de théorie juridique,* 95, 17-64.

Zolo, Danilo (2006), *La justicia de los vencedores. De Nuremberg a Bagdad, Madrid.* Trotta, Madrid.

CAPÍTULO X

INTELIGENCIAS ARTIFICIALES Y LIBERTAD RELIGIOSA: MÁS ALLÁ DE LA DISTOPÍA. UNA PROPUESTA IUSFILOSÓFICA

RAMÓN DARÍO VALDIVIA JIMÉNEZ
CEU Cardenal Spínola. Fundación San Pablo Andalucía
ramvg1974@gmail.com

1. INTRODUCCIÓN

En las primeras horas del 21 de abril del año 2000, un acontecimiento produjo un cambio sustancial en la forma de comprender una de las expresiones de culto público más relevantes del cristianismo: la Semana Santa de Sevilla. La programación de unas carreras descontroladas por parte de unos *amateurs* provocó una avalancha humana que generó violencia, caos y angustia, porque en aquella marabunta se hablaba de que se habían escuchado disparos o agresiones con armas blancas. Días después, la Policía argumentó que el origen del desconcierto fue seguir un "juego de rol". Aún no había tenido lugar la tragedia del 11-S neoyorkino y, sin embargo, ya se percibía la frágil consistencia de la seguridad en una sociedad que tiene como uno de sus factores identificadores las fiestas de contenido religioso.

Si hasta ese momento, las corporaciones religiosas (Hermandades y Cofradías) habían auto-gestionado la organización y el desarrollo de ese tipo de fiesta casi en exclusividad, una de las consecuencias inmediatas de aquel acto vandálico fue el desarrollo exponencial de la implantación de un órgano multidisciplinar de control que pudiera proteger esas fiestas de las situaciones de riesgo. Así se sustituyó el protagonismo organizativo de esas corporaciones religiosas por otras de carácter administrativo y policial (denominado CECOP), teniendo, a partir de entonces, aquellas instituciones un papel simplemente subsidiario (Abascal, 2016, 29). A pesar de la eficacia de este instrumento administrativo-policial, las avalanchas descontroladas en la Semana Santa se reprodujeron en los años 2004, 2009, 2015 y 2017, demostrando que la iniciativa delictiva tiene un carácter anti-religioso. Con el paso del tiempo, dicho instrumento policial, CECOP, se vio abocado a servirse de las nuevas tecnologías, en las que últimamente destaca la Inteligencia artificial, la robótica, el *big data* y las tecnologías de las *Smart cities,* con las que se espera que el control ciudadano pueda ser más efectivo. De esta forma, podemos conectar el mundo de la Inteligencia artificial y su influencia en el ámbito de un Derecho fundamental: la libertad religiosa.

Como he presentado, la planificación estratégica de esa fiesta, que además del contenido religioso adquiere una relevancia económica incuestionable, antaño se resolvía mediante modelos de toma de decisiones protagonizados exclusivamente por esas instituciones religiosas, reconocidas con personalidad jurídica. Pero, en muy poco tiempo, ha pasado a regirse por modelos de inteligencia artificial digital, mediante programas software que interactúan con usuarios que, al realizar tareas tan repetitivas como las de control (López, 2018, 55), pueden llegar a afectar el derecho subjetivo de libertad religiosa cuando: a) se registren en esos programas determinadas imágenes o datos de individuos que no han expresado su consentimiento para ser controlados y b) cuando, deliberadamente, por medio de esta tecnología, la fiesta de contenido religioso pueda llegar a modificarse sustancialmente por los criterios emanados de una IA programada con sesgos anti-religiosos. En este segundo caso, obviamente, vulneraría los principios y valores del ordenamiento jurídico europeo y los derechos fundamentales que se consagran en la Carta de Derechos fundamentales de la Unión Europea, tal y como advierte Llano Alonso: «En un contexto democrático como el de los países de la Unión Europea, es necesario que el Derecho de los robots esté en consonancia con los principios y valores del ordenamiento jurídico europeo» (Llano, 2021, 225).

Aparentemente esta segunda posibilidad puede resultar una hipótesis tan extrema que auspicie todo tipo de conjeturas, como refiere Pérez Luño (2021, 43), pero, a fuer de ser calificado como catastrofista, precisamente, por el hecho de la creciente tendencia hacia un control exhaustivo de la administración en vida social que afecta también en el entorno digital, la jurisprudencia del Tribunal de la Unión Europea se ha planteado la posibilidad de implantar mecanismos garantistas de la libertad religiosa. Por ejemplo, a través de una autoridad de control independiente de los datos de carácter personal, cedidos a la institución religiosa, véase el art. 91.2 del Reglamento 2016/679 de 27 de abril de 2016 relativo a la protección de las personas físicas en lo relativo a los datos y su circulación (Ulloa, 2018, 46). En mi opinión, en el caso de la manifestación pública de la fe que he comentado, el problema no reside tanto en la cesión y control de los derechos sobre los datos de los colectivos religiosos, como del control sobre quienes no han estipulado dicha cesión, pero sus datos sí han sido recogidos en esa manifestación, y puedan ser usados sin su consentimiento.

En el contexto de esos eventos de manifestación pública de la fe, el acceso a los datos personales de quienes participan es relativamente sencillo, especialmente a través de capturas de pantallas de televisión. En situaciones de normalidad, en esos medios de control social parece que el derecho a la libertad religiosa está más que garantizado. Pero y ¿cuándo no se den esas circunstancias, y estemos en un momento de excepción? ¿Puede la

administración policial tratar esos datos limitando la esfera personal protegida por el derecho a la libertad religiosa? La respuesta a esta pregunta parece encontrarse en el art. 5 de la Propuesta de Reglamento del Parlamento Europeo sobre normas armonizadas en materia de Inteligencia artificial, en el que se prohíbe la puesta en servicio y comercialización de sistemas de IA que puedan provocar perjuicios físicos o psicológicos [art. 5.1 a] De modo que, la propia regulación advierte que, sobre el caso específico de la información biométrica remota "en tiempo real", en espacios de acceso público, la comercialización de estos sistemas que pueden hacer una búsqueda selectiva de identidades a través de los rostros puede suponer una amenaza, tal y como se advierte la Comisión Europea (Propuesta COM (2021) 206 final, 6). En efecto, el elemento distópico al que nos referimos en este capítulo aparece cuando se conjuga el flujo del comercio con los datos sensibles de las personas físicas. Lo que induce a pensar que, en el futuro, una Inteligencia Artificial fuerte podría llegar a convertirse en un instrumento que la tradición iusfilosófica ha denominado «tiránico».

Antes estos riesgos ¿pueden asumir estos instrumentos de Inteligencia Artificial estas responsabilidades administrativas?, es decir, ¿puede la administración pública servirse de modo indiscriminado de ellas?, o, por el contrario, vincular a la Inteligencia Artificial el riesgo de la manipulación ¿es la excusa perfecta para que no pueda/ no deba ser utilizada? En estas preguntas subyace la controversia ética sobre el reconocimiento de la personalidad jurídica (capacidad de obrar por sí mismos, y por tanto, de ser responsables de sus acciones) de determinados mecanismos electrónicos independientes de la responsabilidad humana. El Parlamento Europeo se ha manifestado al efecto en la Resolución de 16 de febrero de 2017, en relación con aspectos del documento de la Comisión *Civil Law Rules on Robotics*, en el que se muestra favorable hacia un cierto reconocimiento de *personalidad electrónica*, siempre y cuando exista un código de conducta para los ingenieros, es decir, para las personas físicas que se puedan hacer responsables de los robots y la inteligencia artificial (Resolución Parlamento Europeo 16 febrero 2017, *Civil Law Rules on Robotics*, nº 11). En contra esta hipótesis se ha pronunciado el magisterio católico en el ámbito europeo, a través del documento *Robotisation of life* desaconsejando la posibilidad de otorgar la personalidad jurídica a estas inteligencias artificiales con el argumento del peligro que supone equipararlas a la persona humana y su dignidad, reafirmando así la primacía de la persona humana como único responsable de las acciones de los robots y de la inteligencia artificial (Comece, 2019, 4). Así pues, la primera de las preguntas que nos planteamos es acerca de esta perspectiva distópica, concretamente, ¿qué puede representar el robot en el imaginario iusfilosófico?

2. ONTOPOLÍTICA DE LA ROBÓTICA E INTELIGENCIAS ARTIFICIALES: ¿CABE PENSAR EN LA TIRANÍA?

Para comprender cómo el fenómeno de un robot o de la capacidad predictiva y organizativa de las llamadas inteligencias disruptivas han podido vincularse a la distopía, debemos centrarnos en cómo, en la literatura moral, el argumento de la capacidad de esta tecnología para poder registrar acciones de los ciudadanos ha generado una gran desconfianza, sobre todo porque la misma administración solicita continuamente datos sensibles a la dignidad humana. El gran problema es que estos datos se han convertido en el material más preciado para el desarrollo de la mercadotecnia. La perspectiva del contenido ontopolítico de estas nuevas tecnologías puede analizarse desde lo que actualmente alcanza, sin que podamos ser alarmistas, sobre todo en el espacio político europeo en el que las últimas intervenciones destacan el compromiso con las garantías de las libertades y derechos humanos; pero también supone un ejercicio de racionalidad iusfilosófica el que estas nuevas tecnologías disruptivas puedan analizarse desde una indagación o ficción teórica, como sostiene Biset (2013, 130).

En línea de principio, entiendo que, lejos de ser una ficción tecnológica abstracta, la capacidad de gestionar, coordinar y hasta comerciar con los datos personales por medio de las inteligencias artificiales adquiere un alcance tan estructuralmente político que, en el caso de que esta capacidad pudiera estar monopolizada, la influencia social que tuviera esa capacidad podría ser identificada a un régimen tiránico, en cuanto que podrían desaparecer las garantías y protecciones para la autonomía de los legisladores, siendo ocupadas por los creadores de códigos informáticos, los cuales, sin que tuvieran necesidad de conocimientos jurídicos o legitimidad política, estarían dictando con su ingeniería técnica las reglas necesarias para el desarrollo de unos sistemas que, al fin y al cabo, servirían para regular la sociedad (Fioriglio, 2021, 126).

Para comenzar, tomamos como referencia la figura del tirano en el pensamiento iusfilosófico como una de las constantes políticas de la concepción del poder. Desde el origen de las narraciones mitológicas, el tirano aparece como un sujeto que pretende ejercer el poder omnímodo en su propio beneficio, en contraste con el héroe, quien se resistía al poder hegemónico del tirano haciendo gala de la épica sobre la libertad individual y del beneficio de la comunidad. Así, el modelo clásico de resistencia lo ostenta Teseo, el príncipe de una Atenas pre-política, quien se convierte en el símbolo del ansia de libertad y desarrollo de las polis frente al Minotauro de *Knossos* quien, aislado en el laberinto del palacio del rey Minos, ostentaba su tiranía subyugando a los ciudadanos de las islas griegas mediante su poder. Sin embargo, el

todopoderoso minotauro resultó vencido por la astucia de Ariadna y la valentía de Teseo, de modo que la fragilidad de la individualidad pudo hacer frente a la soberbia talasocracia del mediterráneo mítico (Páez, 2011, 19).

No somos dioses, pero sí humanos capaces de crear máquinas que otorgan fuerza, poder la capacidad de hacer el bien y también de destruir. En el Monte del Olimpo se rumorea que los humanos estamos creando máquinas más inteligentes que nosotros mismos. Solo han pasado unos pocos milenios desde que Zeus actuase tan vilmente. Parece evidente que los mismos temores que Zeus albergó contra Atenea ahora se reencarnan en máquinas que ostentan una inteligencia artificial (Latorre, 2018, 18).

Otras analogías políticas, que en este sentido se han representado en el pensamiento iusfilosófico, acerca de las distintas formas de gobierno y los límites del poder, han partido del sutil análisis de Aristóteles sobre la figura del tirano, como la de aquel que se sirve de técnicas para alimentar la desconfianza y la delación entre los ciudadanos para hacerlos incapaces de operar políticamente (Aristóteles, 1988, 350), algo de lo que continuamente se acusa a estas tecnologías. En efecto, no es poco común percibir cómo las grandes empresas tecnológicas se sirven de los datos que los ciudadanos les proporcionan libremente, para usar sus servicios de redes sociales con los que se ha constituido un verdadero conglomerado de asociaciones mercantiles que se han convertido en imprescindibles para la vida humana, razón principal que se arguye en la sentencia de la Corte General de la Unión Europea contra Google (STJUE, 13 de mayo de 2014 Google Spain c/ AEPD Asunto C-131/12), por ejemplo.

También en Roma, Cicerón presentó la tiranía como la imagen de la concentración del poder, que desfigura el vínculo jurídico (*ratio iuris*) de una comunidad política en el arbitrio irracional de un solo individuo, convirtiéndola en una multitud informe, privada de todo derecho político y civil (Cicerón, 2014, III, 23), como se ha percibido en la demanda de la Comisión Federal de Comercio de los Estados Unidos contra Facebook, instándole a vender Instagram o WhatsApp a Mark Zuckerberg.

Un cuarto momento fundamental, que precisó el concepto de tiranía, fue la cultura cristiana representada por Tomás de Aquino, quien precisó que debía aplicarse ese título a quien abusara del indiscriminado uso de la violencia como método de adquisición del poder (Tomás de Aquino, I - II, q. 105, art. 1 ad 2). Desde esta perspectiva, no me encuentro capacitado para vincular la tesis tiránica a la Inteligencia Artificial por este medio, sin embargo, la génesis de la tesis del poder disruptivo de las nuevas tecnologías vincula la importancia de la Inteligencia Artificial con la carrera armamentística, especialmente por su capacidad de gestionar y coordinar datos de población, territorios, etc.

Por último, a las puertas de la modernidad iusfilosófica, desde la racionalidad ético-política tenemos dos percepciones del tirano: la primera, le permitió a Bodino diferenciar las figuras de la tiranía del despotismo, porque la primera despreciaba el derecho natural a la libertad y a la propiedad privada de los ciudadanos (Turchetti, 2008, 36), tal y como puede expresar la IA desde la administración de los países de la esfera comunista, que entiende que los datos pertenecen al Estado, tal y como sucede en la República de China; y desde la segunda perspectiva, a Hobbes le atraía la comprensión de la tiranía como una necesidad social, porque permitía evitar el mal absoluto de la guerra civil que promovían las religiones, proporcionando así el valor de la seguridad jurídica como valor supremo, y doblegaba otras exigencias fundamentales, como la de la libertad religiosa, aunque fuese a costa de las incomodidades políticas que suponía aceptar dicha tiranía, como pudiera ser precisamente el caso de que se pudieran vulnerar los derechos y libertades fundamentales, como la no discriminación por razón de raza, sexo o religión para salvaguardar una exigencia de seguridad (Bubner, 2015, 219).

En efecto, como conclusión de este somero estudio sobre la *ontopolítica* del tirano, defino su figura como la técnica-política que desarticula la capacidad política de la comunidad a la que dice representar, desvinculándole los lazos de participación comunitaria porque decide asumir irremisiblemente el gobierno de toda la realidad a cambio de ofrecer seguridad. Y, precisamente, la seguridad es la principal virtud social que puede aportar la IA, gracias a su capacidad programadora.

Desde esta perspectiva, como la robótica tiende a la simple emulación de la humanidad (Pinto, 2020, 107), en vez de ofrecer las pautas que le permitan escapar del determinismo causal, como señala Yuk Hui (2020, 77), es necesario que pueda moderarse la técnica disruptiva que sostiene la IA para que no pueda llegar a convertirse en un instrumento al servicio de un único control del mundo globalizado, tal y como han advertido algunos autores desde una perspectiva catastrofista, entre los que destacamos a Hans Jonas, Habermas o Alexandre. En efecto, Jonas identifica la IA o la robótica con la tiranía por la fractura entre la promesa y la amenaza con la que la robótica puede alterar el orden cosmológico Jonas (2004, 16-17); Habermas, por su parte, por la exigencia de parámetros conservacionistas del debate biológico Habermas (2017, 59) y Alexandre, por la dimensión económica que imprime la cultura tecnológica, ya que invierta ingentes cantidades de dinero para demostrar el abismo cultural que introducen estas tecnologías disruptivas respecto a las capacidades humanas (Alexandre, 2018, 98).

Desde otra perspectiva mucho más moderada, Llano quiere advertir del riesgo gnoseológico que supone la simple ilusión de que esos robots puedan

alcanzar siquiera el nivel de la inteligencia humana (Llano, 2018, 98). En este sentido cultural, propio de la tradición humanista de la identidad europea, heredera del respeto a la integridad del ser humano junto al orden jurídico liberal establecido (y, por tanto opuesto al marco de la tiranía), también el *Libro Blanco de la Inteligencia Artificial* advierte de los peligros que pueden desencadenar el uso de la Inteligencia Artificial más opaca, es decir, aquella que pueda adquirir un comportamiento parcialmente autónomo, entre otros, el de discriminación por razón religiosa que: «pueden ser resultado de defectos de diseño general en los sistemas de IA (…) o del uso de datos que pueden ser sesgados sin una corrección previa (…) [y, en el caso de sea una IA que "aprenda"], a las repercusiones prácticas de las correlaciones o de los modelos que reconozca el sistema en un gran conjunto de datos» (DOC COM (2020) 65 final, pp. 13-14), de modo que, a manera de ejemplos, advierte contra el uso que los estados puedan hacer de ella para la vigilancia masiva o abusar de los datos que consigan.

El realismo jurídico-práctico de la Unión Europea, indica que, a pesar de estas voces críticas de la IA por los hipotéticos sesgos tiránicos, esta nueva realidad disruptiva genera ya unos grandes beneficios para la convivencia y sostenibilidad de nuestras sociedades.

3. EL PRINCIPIO DE PRECAUCIÓN ANTE EL DERECHO A LA LIBERTAD RELIGIOSA EN LA ERA DIGITAL

Sin embargo, a pesar de esos beneficios que ya compartimos en las sociedades en las que está implantada la IA, la estela de Jonas o Habermas lleva a preguntarme por la oportunidad de recuperar unos principios éticos y jurídicos que salvaguarden la humanidad de una posible deriva excesivamente controladora de la IA en materia de religión. Pienso que estos principios deben servir para que la promoción y garantía del derecho a la libertad religiosa no sea una mera entelequia en el enjambre de la época digital, y que los continuos cambios que genera la tecnología no puedan eliminar este derecho fundamental, sustituyéndolo por otros que, sobre la exigencia de seguridad, descarten la pluralidad de conciencia o de convicciones ético-jurídicas reconocida como un principio fundamental o valores material básico del ordenamiento jurídico español (García de Enterría, 2005, 85).

En efecto, ante el vertiginoso cambio que la tecnología ofrece, la dimensión normativa del Derecho, basada fundamentalmente en la ley, parece que es incapaz de alcanzar a regular todos los particulares que puedan generar los avances tecnológicos que puedan estar en contradicción con esas libertades ya conquistadas; por no hablar de la imposibilidad de que fuese la costumbre, como fuente subsidiaria, la que pudiera llegar a regular una cuestión tan novedosa como la que estamos tratando. Del mismo pensar es Piñar Mañas,

quien también recurre a los principios para evitar la obsolescencia del Derecho, desbordada por la evolución tecnológica (2018, 18); o Barrio Andrés, quien subraya que la oportunidad de los principios proviene de la capacidad de inspirar a los poderes públicos y de iluminar sus políticas; o también López Oneto quien, siguiendo la estela de Alexy, señala que: «de la circunstancia de que las leyes (o principios) fundamentales del DIA [Derecho de la Inteligencia Artificial] no consten expresamente los enunciados normativos de la Declaración Internacional de los Derechos Humanos, no se sigue necesariamente que no existan, toda vez que implícitamente puedan estar contenidos en las diversas piezas del Derecho Internacional de los Derechos Humanos» (López, 2020, 201-202).

Entre esos principios, el de seguridad pública fue aducido tradicionalmente en los regímenes tiránicos para sortear cualquier obstáculo en el desarrollo de su poder, y también hoy, al recurrir a este principio para la defensa de los derechos y libertades fundamentales, se pueden infiltrar sesgos ideológicos en los mecanismos de control de la administración pública o en las compañías privadas para la selección de personal, por ejemplo. Estos sesgos pueden alterar la naturaleza del resto de principios con los que debe convivir la seguridad, como es el caso de la libertad religiosa, de modo que ya no sea suficiente la mera programación legalizada según las exigencias de la dignidad humana, sino que sea necesario además revisar conceptos que, por estar precisamente asumidos en el acervo jurídico, puedan ser pasados por alto y como consecuencia, sean alterados en su función ética, tales como el lazo social o el cuidado común como funciones del Derecho (Ferreira, 2021, 109).

3.1. Principio de Precaución

Por ello, me parece interesante aplicar, mediante la *analogia iuris*, el principio de precaución al uso de la robótica, el *big data* y la IA. Ya comenzó su uso en casos de materia medioambiental, pero se ha extendido sobre la protección de otros bienes jurídicos que afectan a la persona humana, imponiéndose sobre otras contraprestaciones, especialmente de índole económica (Cierco, 2004, 94). Este principio, reconocido como principio general del Derecho comunitario, se ha ido expandiendo en el Derecho internacional muy lentamente, y se aplica cuando: «existe una apreciable incertidumbre científica acerca de la causalidad, la magnitud, la probabilidad y la naturaleza del daño» (López, 2020, 158), de manera que es en este estado de incertidumbre cuando aparece que no debe hacerse uso de estos métodos disruptivos para garantizar la seguridad, aunque se pretenda contener un daño, porque su uso puede llegar a vulnerar derechos fundamentales como el de libertad religiosa.

De este modo, entra en juego el conflicto entre la seguridad que ofrece la IA, como un método que prevé los riesgos, y el hecho de que no haya apenas

espacio humano en el que no se haya expandido la inteligencia artificial. Por eso, la propia ONU está exigiendo regulación directamente proporcional al desarrollo de la IA, pues no se trata de limitar su desarrollo, en detrimento del propio hombre, sino de ofrecer cauces éticos para el bien de la humanidad (Zhao, 2018). En efecto, la exigencia ética del principio de precaución, que ya estaba previsto en la literatura de ficción de Isaac Asimov cuando decía: «un robot no hará daño a un ser humano, ni permitirá que, por inacción, el humano sufra daño» (Asimov, 1942), se ha convertido en un principio jurídico al tener como contenido primordial la dignidad humana sobre el avance científico o tecnológico, como expresan tanto en la legislación española (art. 10.1 CE) como los convenios internacionales (BOE 251, de 20 de octubre de 1999 art. 2).

Un ejemplo de este principio de precaución, derivado de la anterior legislación supone que, en aras del respeto a la dignidad humana, en todos los procesos bioéticos que suponga el uso de un robot, una persona humana debe ser el verdadero interlocutor de la administración pública ante los pacientes, si se quiere respetar los derechos fundamentales, haciéndola a esta responsable civil, penal o administrativa de los daños ocasionados por el robot. En este sentido, además, otras instancias que la legislación europea ha tomado como referencia para el estudio de la IA, como el informe del Instituto Reteneu de 2017 "*Human Rights in the robot age*", advierte que debe tenerse en cuenta el principio de precaución, efectivamente, en la necesidad de protección de los datos personales y de la privacidad que requiere el ejercicio de la conciencia personal y de la libertad religiosa, en especial en la exigencia de determinadas plataformas masivas de datos que contienen empresas privadas como Facebook (Van Est, Rinie, Joost Gerritsen, 2017, 18).

3.2. **Precaución como salvaguarda de la libertad religiosa**

En el caso que presento de las avalanchas programadas para interrumpir el sereno desarrollo de una actividad religiosa en el espacio público, la indefensión puede proceder del estudio y publicación de los datos obtenidos por las cámaras televisivas sin previo consentimiento; la publicación de esos documentos en medios de comunicación; la limitación de circulación en espacios públicos; las prohibiciones de manifestaciones públicas de la experiencia religiosa; o incluso, la demostración de expresiones de odio sobre estas manifestaciones religiosas. Todo ello atentaría contra los derechos fundamentales, tan protegidos como la misma seguridad que invocan los administradores de esos instrumentos tecnológicos, como demuestra la Declaración Universal de los Derechos Humanos de 10 de diciembre de 1948 que protege la libertad individual como valor primario (art. 1); la no injerencia o arbitrariedad respecto a la honra o reputación (art. 12); o la libre circulación (art. 13). Una libertad que, está indisolublemente ligada a la exigencia de seguridad

de los convivientes (DOUE 83, de 30 marzo de 2010, art. 2), de modo que los legisladores internacionales han solicitado que estos límites deban ser regulados por ley. Pero, de nuevo, aparece el problema de la obsolescencia legal frente al avance de la tecnología disruptiva.

De forma que, en este caso, que afecta a la libertad religiosa, parece necesario, según el propio el propio Instituto Reteneu, que el pretendido desarrollo de esta tecnología esté siempre bajo la supervisión y responsabilidad de una persona humana, especialmente en el ámbito de la toma de decisiones, con lo que el denominado principio de precaución devuelve el protagonismo y responsabilidad a aquella persona humana que ha podido presentar los datos a la Inteligencia Artificial para que pueda analizarlos, sin que sea esta la que deba asumir la plenitud decisiva en las situaciones límite.

Como señala Barrios Andrés, en tal caso: «Corresponderá entonces a la sociedad en su conjunto predefinir cuidadosamente lo que está dispuesta a tolerar para que este nuevo instrumento pueda utilizarse, o no, en circunstancias en las que pueda atentar contra las libertades constitucionales de las personas» (Barrio, 2019, 239). El problema, a mi juicio, reside en que parece que la sociedad puede estar tan anestesiada ante la evolución de esta tecnología que no encuentra frenos ni resortes que permitan controlar la evolución de la IA para que llegue tomar conciencia del daño que puede llegar a provocar, pues parece que toda limitación al desarrollo tecnológico o la libertad de expresión en las mismas redes sociales se percibe de forma negativa, tal y como se hacen eco tanto los organismos internacionales (Guterres, 2019), como en las instancias encargadas de vigilar la legalidad respecto a la incidencia de los delitos de odio por motivos religiosos en las redes sociales (Circular 7/2019 de 14 de mayo, de la Fiscalía General del Estado sobre pautas para interpretar los delitos de odio tipificados en el art. 510 del Código Penal, en BOE 124, de 24 de mayo de 2019, p. 14)

Sin embargo, algo mucho más sutil, y nefasto para el ejercicio de la libertad religiosa sería que, quien modulara el sistema de los algoritmos del robot o del complejo sistema de la Inteligencia Artificial y *big data* pudiera llegar a determinar por sí mismo qué tipo de expresión pública sea para la sociedad tolerable, volviendo como en un *deja vu*, al momento de la constitución de los principios lockeanos de la democracia liberal. Si llegara este momento, entonces se podría entrar en otra encrucijada perversa acerca de la cuestión religiosa, una diatriba fatal entre la posibilidad de optar porque sea el Estado el que legisle de manera exclusiva sin atenerse al respeto de la conciencia individual, o bien, la posibilidad de que el mismo Estado abandonase su capacidad legislativa, y reconociera su impotencia ante el desafío tecnológico, por más que tenga que

reconocer que esta "condicionado" por la obsolescencia legislativa (Navarro-Valls/Martínez Torrón, 1997, 246).

3.3. **Principios en falso**

Para solventar esta diatriba, la ideología *globalista*, que pretende acomodar las diferencias religiosas, morales y metafísicas en el conjunto de los Derechos Humanos ya existentes (Maffetone, 2019, 28), ha recurrido al desarrollo de la cultura ética global, de amplias resonancias kantianas, acerca de los márgenes racionales de la religiosidad, con la que se procura un consenso elaborado que limita el credo religioso, en beneficio de una actividad ética derivada de la conciencia religiosa, cuyo contenido se puede especificar en torno a los grandes temas en los que la convergencia ética pudiera construirse con mayor facilidad, debido a su capacidad para generar una cultura de la paz, siempre que se sustente sobre los valores del amor, la alegría, el dolor, la felicidad, el destino, la justicia, y en último término, la vida y la muerte, y que se obvien las graves diferencias de contenido dogmático entre ellas (Ünver, 2018, 256). De esta manera, la distopía que "denunciaba" Locke en sus principios proto-liberales, parece solventarse en favor de una ética global, de modo que los asuntos que plantean conflicto religioso, considerados por Locke como *adiaphora* o irrelevantes, se oscurecen en beneficio del bien común (Locke, 1999 [1661]). Así pues, según la ética inspirada por este movimiento *globalista*, la instancia ética que debe decidir el contenido de las expresiones religiosas debe someterse a los principios racionales, promovidos desde su propia perspectiva, la cual puede ser diseñar y programada según el complejo sistema del *big data*. Y, conforme a estos compromisos universales, puede llegar a prohibir, o censurar, aquellas expresiones religiosas que no se adapten dentro de estos márgenes de racionalidad, impidiendo su difusión, por ejemplo, en las redes sociales, en el ámbito privado, o en los medios de comunicación de dominio público, si es que afectan a la Administración.

3.4. **El tratamiento de los sesgos contra la discriminación religiosa: identificadores de sesgos**

En efecto, si dejamos todo el control de lo permitido a las conciencias religiosas en manos de un procedimiento de IA, que respete exclusivamente el ámbito de esa ética religiosa *globalista*, y que pueda estar sujeta además a los sesgos culturales de un determinado grupo de poder, ¿no puede resultar sospechoso este procedimiento de atentar contra el principio de libertad religiosa? Considero que el desarrollo de la IA parece inevitable, ahora bien, podemos articularla siguiendo el mismo procedimiento con el que se estableció el consenso democrático español, registrado tanto en la CE como en la LOLR, ya que de seguir la lógica reductiva de la racionalidad moderna sobre la libertad religiosa, la IA podría ser un instrumento exorbitante, usado por quienes

simplemente o la censuren o bien impongan el criterio de la ética religiosa *globalista*, por ejemplo, a través de algoritmos que contengan sesgos contrarios a la libertad religiosa.

Sobre cuáles puedan ser esos sesgos, a pesar de la presión que ejerce la obsolescencia del Derecho, se ha pronunciado la Organización para la Seguridad y Cooperación en Europa (OSCE), que ha denunciado la discriminación o vulneración del derecho a la libertad religiosa por la utilización de determinados *sesgos identitarios* en la cultura contemporánea, aunque no de manera explícita para el entorno digital. En efecto, este organismo internacional ha articulado unos principios de salvaguarda del derecho fundamental a la libertad religiosa que sirven como un vehículo probatorio de estas conductas antidiscriminatorias a través de unos *indicadores de sesgo,* con una especial implantación respecto el antisemitismo, y que pretende servir de criterio objetivo para una imputación válida, y análogamente, a otras experiencias religiosas (OSCE, 2017, 50). Considero que estos *indicadores de sesgos* deberían introducirse en la información para el algoritmo que regule la convivencia en ámbitos donde pueda proceder el uso/abuso de la experiencia religiosa. La OSCE advierte, como no, de la percepción no sólo de la víctima, sino también de testigos o de peritos que pudieran estar condicionados, por eso, dice que debe atenderse principalmente a los comentarios, declaraciones escritas, gestos o grafitis que estén relacionados con conductas que vulneran la libertad religiosa, especialmente reguladas en expresiones antisemitas o con un contenido enaltecedor de la supremacía nazi.

También la OSCE ha desarrollado un documento sobre la protección a la libertad religiosa islámica, y aunque tenga un menor contenido formal, se sobreentiende que los principios que protegen a la comunidad judía deben ser válidos también para la islámica, de modo que lo que aparece es sólo la clarificación acerca de los sesgos discriminatorios que afectan a la comunidad islámica (OSCE, 2018). Me resulta muy interesante, por otra parte, cómo en el ámbito de la OSCE, la protección jurídica y social se ha dirigido principalmente sobre las religiones minoritarias de mayor arraigo, tanto la religión judía como la islámica (RD 593/2015 de 3 de julio, art. 3), mientras que la religión mayoritaria, la cristiana católica, aún no tiene ningún documento marco con el que pueda desarrollarse una futurible regulación de indicadores de sesgos.

La OSCE deduce estos indicadores de sesgos a través de preguntas explícitas o implícitas que revelan no sólo una determinada ideología, sino el nivel de compromiso del sujeto que atenta contra la libertad religiosa. A través de unas preguntas se revela como un indicador objetivo acerca de la pertenencia a grupos de delincuencia organizada, ya sean de alcance internacional o de pandillas que estén motivadas por el odio. Otras referencias

que pueden insertarse en la información del gestor IA refieren al espacio y el tiempo en que se produjeren los delitos de odio, tales como espacios y fechas simbólicas: Palestina, Estado de Israel, barrios de mayoría judía, cementerios judíos o fechas que expresen un contenido religioso, como el Yom kipur, Pésaj, Ros Hashaná, o el de sus victimarios, como las fechas representativas de altos dirigentes nazis, o de acontecimientos ligados a la historia antisemita.

De esta manera la introducción de estos sesgos nos parece de singular importancia para la aplicación y eficacia de la IA por la capacidad de relacionar acontecimientos, espacios y tiempos es la capacidad para dibujar los patrones delictivos con los que se reflejen la naturaleza de estos delitos de violencia. Por eso, considero que, además de la elaboración internacional, o estatal, de estos indicadores, también pueden resultar muy útiles algunos observatorios u oficinas especializadas que estén atentas a las vulneraciones del derecho fundamental sobre libertad religiosa, tales como los que han propiciado algunas comunidades religiosas, que puedan ayudar a eludir la distopía de un control irracional por parte de quien elabora la IA.

4. UNA PROPUESTA: COLABORACIÓN PARA ELUDIR LA DISTOPÍA

Mi propuesta para vencer la imagen distópica que puede generar la IA en el ámbito de la libertad religiosa consiste en facilitar acuerdos de colaboración entre las diferentes comunidades religiosas, la Administración pública, y por supuesto, los programadores de los algoritmos. Este hipotético consenso padecerá, como el propio Derecho, las limitaciones temporales de los acuerdos, sin embargo, nada obsta a que pueda ser posible dados los antecedentes que se han producido ante los avances en otro tipo de delitos de la esfera religiosa, como el de los delitos de odio. Así, mi propuesta supone una hipótesis de *analogía legis* sobre esta figura, para aplicarla a la regulación de los supuestos de vulneración de la libertad religiosa por parte de una Inteligencia Artificial que haya adquirido características demasiado fuertes, o al menos, haya sido programada vulnerando la libertad religiosa.

Aunque entre las recomendaciones que la OSCE y la Oficina de Instituciones Democráticas y Derechos Humanos (ODIHR) aparezca la necesidad de capacitar a los agentes de policía para que puedan centrarse en los métodos que identifiquen los delitos de odio, así como en la formación de habilidades para compartir inteligencia y trabajar con fiscales y las comunidades afectadas (OSCE, 2009, 60), quienes hoy están realizando un estudio de campo que analiza este problema de dignidad de la persona, son las propias comunidades religiosas. Especialmente desarrollada están las comunidades judías, tal y como refleja la guía sobre los delitos de odio de naturaleza antisemita, que tomamos como referencia para nuestra propuesta. En efecto, en la guía que desarrollaron estas comunidades junto con la OSCE y

la ODIRH, se expresa la necesidad de: «Facilitar el intercambio de buenas prácticas entre los Estados participantes en la OSCE, con particular atención a los modelos de colaboración entre las comunidades judías y las fuerzas policiales» (OSCE, 2017, 3). El mismo documento de la OSCE reconoce la importancia de la colaboración institucional entre la policía y las comunidades religiosas, y del uso de la inteligencia (y, aunque no lo exprese literalmente, se entiende que sea necesaria también *el uso* de la inteligencia artificial) para frenar los abusos contra la libertad religiosa del siguiente caso:

En la República Checa, en respuesta a una manifestación neonazi que iba a atravesar el barrio judío de Praga para conmemorar el pogromo de 1938 contra la población judía en la Alemania Nazi (*Pogromnacht*, también conocida como la «noche de los cristales rotos»), las comunidades judías, en estrecha cooperación, comunicación y coordinación con las autoridades checas, adoptaron una amplia serie de medidas de seguridad con carácter previo y durante la manifestación. […] Esta mayor colaboración también requirió el establecimiento, puesta a punto y uso de una sala de control, reuniones informativas conjuntas, intercambio de inteligencia antes y durante el evento, el levantamiento de barreras y cierre de carreteras, así como el establecimiento y control conjunto de puntos de control. Este esfuerzo mutuo se tradujo en una oportunidad para generar confianza entre ambas partes y evitar la duplicación de esfuerzos. (OSCE, 2017, 37).

En el mismo documento, se reconoce claramente que la aportación de información fidedigna acerca de los delitos de odio resulta «primordial para que los gobiernos puedan evaluar los datos planteados por el antisemitismo, [y que] no recabar dicha información podría percibirse como un intento de minimizar la importancia del problema o negar su existencia» (OSCE, 2017, 43-44), con lo que deja que el punto de partida de esta información proceda de las víctimas, alentadas por las mismas comunidades religiosas, como mediadoras o cauces de la denuncia, para que sean los instrumentos del Estado los que posteriormente desarrollen el proceso oportuno, por ejemplo, en el caso del análisis de las pruebas (De Asís, 2021, 81). Este protagonismo de la sociedad civil frente a los delitos de odio refleja la necesidad de que los delitos de odio u otros que vulneren la dignidad de cualquier persona, y la dimensión religiosa de la persona es uno de ellos, no pueda ser tolerado bajo la mera ironía, sarcasmo o simplemente la libertad de expresión, pues lejos de percibirse en esta sociedad como un problema serio, en realidad supone una realidad irrefutable.

Al realizar un análisis de este problema en las distintas religiones con mayor arraigo, podemos concluir que la comunidad católica en España no tiene ninguna referencia corporativa para el diálogo institucional sobre estos delitos

contra la libertad religiosa, salvo la salvaguarda que pueda realizar dicha Comisión Asesora.

La mayor preocupación reside en el caso de las comunidades islámicas porque, además de no tener un órgano de vigilancia, su lugar natural de socialización es la mezquita, la cual se ha percibido tradicionalmente como una *amenaza*, tal y como refleja la misma policía autónoma del País Vasco, la cual comunicaba que este lugar era prioritario para la prevención de la radicalización, es decir, lo consideraba con un sesgo delictivo, del que se sirven para el control de los poderes del Estado (Echániz, 2017).

En cambio, la experiencia de la comunidad judía es la más innovadora, y nos sirve como punto de partida para alimentar la necesidad de cooperación entre las comunidades religiosas y los organismos públicos, para armonizar, intercambiar, controlar y recopilar la información de los delitos contra la libertad religiosa. Sobre todo, porque a través de esta colaboración que presta la misma comunidad, la víctima resulta protegida en los procedimientos y denuncias, siempre que se realicen mediante protocolos consensuados en los que se garantice el anonimato de la víctima, pero también su identidad en el caso de que pueda servir de cauce para el abuso de derecho. En efecto, dicha protocolización no puede servir para que las supuestas víctimas puedan adquirir el privilegio que, en principio pueden tener los instrumentos de la IA, como puede ser el anonimato. Además, este documento nos muestra la importancia de que estas comunidades religiosas puedan organizarse para ofrecer un cauce institucional que sirva de instrumento con la administración, pues de otro modo, la relevancia de la información podrá quedar siempre en entredicho en el lado de los poderes públicos. Pone como ejemplo el documento de la OSCE al SPCJ francés (Servicio de Seguridad de la Comunidad Judía [en francés]) con el que trabaja de manera estrecha el Ministerio del Interior por medio de reuniones mensuales. Pero además de servir de estímulo para la denuncia e información de los delitos de odio, dicha comunidad judía puede mejorar la comprensión de los funcionarios públicos en torno a la experiencia religiosa que se ha visto comprometida con el delito de odio, y también el poder ofrecer un cauce de comunicación desde los servicios de seguridad públicos, de manera que sirva para tranquilizar a los fieles en caso de multiplicación de delitos, reduciendo así la tensión social. Sin duda, esta colaboración institucional es un verdadero cauce de paz social que puede ser válido y eficaz también en los delitos que puedan realizarse por medio de la Inteligencia artificial, o bien, a la hora de realizar los mismos algoritmos que produzcan los poderes estatales, e incluso las entidades privadas para evitar, además, la discriminación.

En España, aparte de la comisión que asesora a las Administraciones Públicas en relación con la aplicación tanto de la Ley Orgánica de Libertad Religiosa, regulada por el RD 932/2013 de 29 de noviembre, que actúa de manera colegiada y cuya función es consultiva (Real Decreto 932/2013 de 29 de noviembre, arts. 1 y 3), la comunidad judía tiene su propio observatorio acerca de los delitos de odio (Federación Comunidades Judías, 2019), lo que garantiza la vigilancia sobre las amenazas externas a su religión, también en el ámbito digital.

Por último, debemos destacar que, debido a esa preocupación, la misma sociedad civil ha tomado el problema de la libertad religiosa en serio, y desde los últimos diez años, una institución sin ánimo de lucro, el *Observatorio de la Libertad Religiosa y de Conciencia* (OLRC), se encarga de difundir y distribuir información sobre las vulneraciones a la libertad religiosa. Desde esta institución, por ejemplo, se advierte que: «a lo largo de 2020 la religión que más ataques ha sufrido a la libertad han sido los cristianos, con 174 ataques, mientras que 12 casos se dirigieron contra los musulmanes y 6 contra los judíos» (OLRC, 2021, 8).

Una vez que hemos mostrado el nivel de colaboración jurídico entre las comunidades religiosas y las administraciones para la prevención de los delitos de odio, me parece razonable que, en un ámbito como el de la inteligencia artificial, también pueda ofrecerse un camino de colaboración entre estas comunidades religiosas, la administración pública e, inevitablemente, las entidades privadas que crean y gestionan los algoritmos para que pueda desarrollarse efectivamente el ejercicio del derecho a la libertad religiosa.

Partiendo del hecho de que sea más que previsible que, con la obsolescencia que caracteriza el derecho sobre la Inteligencia Artificial, no se pueda llegar a regular plenamente los delitos que limiten el ejercicio de la libertad religiosa en el entorno digital. Sin embargo, me parece que, si las mismas comunidades afectadas centraran su atención en esta nueva realidad disruptiva, junto a la colaboración institucional, la colaboración puede ser un freno eficaz a la expansión delictiva a la que puede llevarnos la distopía en un mundo digital. Precisamente el valor de la colaboración o de participación para la protección y el respeto en la autonomía de la gestión de sus datos, incluso entre grupos marginados, ha sido puesto de relieve en el reciente proyecto de texto de la recomendación sobre la ética de la inteligencia artificial de la UNESCO (2021, 25).

El valor de la dignidad humana que, a pesar de los conflictos, se ha constatado en las religiones tradicionales, puede ser alterado por el de las tecnologías disruptivas en cuyas propuestas el hombre pueda ser manipulado, o tratado como un medio por los poderes económicos que sostienen los

creadores de los algoritmos (Pinto, 2020, 194), los cuales tendrán que seguir pautas de transparencia cada vez más sofisticadas en la creación de sus códigos, como sucede también en el ámbito sanitario o en el caso de la seguridad y orden público (Fioriglio, 2021, 129). También el documento de la UNESCO refiere a la religiosidad humana como signo de preferente cuidado y protección de la dignidad humana ante los abusos que pueden proceder de la IA (Unesco, 2021, 20), alertando sobre todo lo relativo a «las medidas de seguridad conexas, los datos biométricos, […] y los datos personales como los relativos a la raza, el color, la ascendencia, el género, la edad, el idioma, [y] la religión» (Unesco, 2021, 31). Pero al mismo tiempo, la UNESCO también ha recalcado el factor positivo de la IA, sobre todo desde una perspectiva inclusiva, en la que no puedan producirse agravios por falta de acceso a las nuevas tecnologías, de manera que puedan distribuirse con equidad los beneficios que ella misma promociona.

En efecto, de la relación entre la administración y las partes interesadas en la IA ya se hablaba en la Comunicación *Inteligencia Artificial para Europa* a la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Religiones de 25 de abril de 2018 proponiendo una colaboración permanente entre los interlocutores sociales, entre los que destacaba las asociaciones entre empresas y centros educativos y administraciones para incluir la IA y su impacto en la economía, el empleo, la diversidad y el equilibrio de género, y también en ofrecer unas directrices éticas en relación con la IA (DOC COM (2018) 237 final, p. 16). Y la Resolución del Parlamento Europeo de 12 de febrero de 2019, sobre una política global europea en materia de Inteligencia artificial y robótica, entiende la utilidad que, sobre el problema de los límites de la autonomía de la inteligencia artificial y la robótica, la colaboración entre la IA y la acción humana es imprescindible (Resolución Parlamento Europeo, de 12 de febrero de 2019, sobre una política industrial global en materia de inteligencia artificial y robótica (2018/2088 (INI), nº 152), si bien requiere la ayuda de autoridades de supervisión independientes que garanticen la transparencia y un tratamiento adecuado de la seguridad jurídica en términos generales que pueda estar dotadas, además, de recursos financieros, como lo presenta en el número siguiente de la misma resolución (nº 159).

Este es el principio que fundamenta el principio del *Libro Blanco sobre la Inteligencia Artificial* de la Comisión europea que, además de pretender generar un *ecosistema de excelencia*, requiere un *ecosistema de confianza*. Esta necesita el respeto a las normas de protección de los derechos fundamentales, especialmente en aquellos sistemas de IA que puedan presentar un riesgo elevado de vulneración de estos derechos. Así, la Comisión europea dice querer respaldar un enfoque decididamente antropocéntrico, en el que la

comunicación de los productores de la IA pueda generar confianza (DOC COM (2020) 65 final, 25), que es el presupuesto opuesto a la distopía. La Comisión europea entiende que la capacidad de innovación en IA puede desarrollarse sin que se tenga que renunciar a los principios antropocéntricos, que quizá no están presentes en el desarrollo de competidores estatales y/o industriales en los que la ausencia de estos principios puede generar los perjuicios propios del descontrol. La seguridad y defensa de la dignidad humana, por el contrario, lejos de suponer una rémora se percibe como un verdadero incentivo para que la investigación, la innovación y el desarrollo puedan generar confianza. En esta línea, además, la Comisión apuesta por la colaboración institucional para consolidar no sólo la excelencia, sino el clima de seguridad jurídica, por eso considero imprescindible que, en esa comunicación puedan estar presentes también esas esas autoridades de supervisión, entre las que pueden estar esos organismos de seguridad de las comunidades religiosas que puedan ofrecer las garantías de un tratamiento adecuado de la IA respecto al derecho fundamental de la libertad religiosa, cuidándose de que no puedan inmiscuirse no sólo actividades delictivas como las de los delitos de odio, control excesivo, o tratamiento ilícito de los datos de naturaleza absolutamente privada porque afectan a la conciencia humana.

Otro tema de interés para el que podría servir esta colaboración sería el control sobre el abuso que alternativas religiosas, que usan la IA como fuente de inspiración religiosa, puedan servir como fraudulentas experiencias pseudo-religiosas.

Finalmente, volviendo al caso inicial que nos ocupaba, la simple voluntad de unos jugadores en un "divertido" juego de rol que generó una situación muy compleja en el año 2000, pudo haberse evitado con la tecnología que hoy disponemos de información con cámaras controladas por medio de transformación de imágenes en datos, y robots capaces de elaborar estos mediante algoritmos capaces de identificar a los promotores de las estampidas. Esta capacidad de la IA de volver transparente la vida de los ciudadanos a través de una recolección indiscriminada de datos, volcados además a través del uso masivo de las redes sociales, aunque pueda llegar a hacer innecesaria la investigación penal, vulnerando el principio de la presunción de inocencia (Nieva, 2018, 150), también puede servir como herramienta para la lucha contra los sesgos y prejuicios (Laukyte, 2021, 195).

Sin embargo, lamentablemente, la preocupación que origina conduce a la siguiente pregunta: ¿quién dice que, dado que se puede hacer perfectamente el seguimiento de multitudes, no se pueda realmente establecer unos patrones en los que el derecho a la intimidad de la conciencia y a la práctica de la religión no puedan ser controlados para fines delictivos? La utilización de los perfiles de

riesgo puede ser usada especialmente en esos ámbitos religiosos en los que, como reconocía la Policía Autónoma Vasca, estaban ligados a la convivencia social de las mezquitas, porque son lugares donde actúan los sesgos religiosos de manera más acentuada que en otras comunidades religiosas, pero la historia nos ha mostrado cómo se pueden tornar las condiciones para que otras comunidades religiosas puedan ser vistas como un peligro público.

5. CONCLUSIONES

En el presente capítulo se ha invocado la importancia de la Inteligencia Artificial como un hecho ya presente en la vida cotidiana de los ciudadanos a través de los robots, en la medicina, en la organización de trabajos, en eficiencia de los recursos, en la sostenibilidad de las ciudades, y también en uno de los aspectos más controvertidos: la vigilancia de los espacios públicos y el control del movimiento de los ciudadanos, especialmente en momentos de crisis.

Precisamente, en este aspecto, y en la posibilidad de que la IA haya podido alimentar una imaginación desbordante sobre la valencia distópica en la literatura ético-moral, la realidad se fragua entre los conflictos que pueden generar las nuevas tecnologías y las exigencias que el marco normativo europeo impone como criterios de defensa de los valores de la dignidad humana y los principios de respeto a los Derechos Humanos. A pesar de la obsolescencia del Derecho, y su muy compleja capacidad para seguir el paso a la evolución tecnológica, esta cuestión, lejos de ser algo olvidado, se está tratando de normalizar, también en el respeto a la libertad de creencias. El Derecho, efectivamente, no puede llegar a controlar el deseo de las grandes corporaciones, o de los gobiernos con sesgos más autoritarios, para obtener los datos de los presentes y futuros clientes, o de sus propios ciudadanos, a través de la IA. Esta tendencia, que la literatura ética ha podido denominar tiránica, ha sido generalmente frenada en su ambición por medio de los instrumentos a disposición del Estado de Derecho.

Percibir la evolución tecnológica y la introducción de la IA en los Cuerpos de Seguridad del Estado, que han optado por el uso de esta revolucionaria capacidad de controlar espacios públicos, masas de gentes y situaciones de crisis, lejos de generar sospechas ha sido reconocido como una ayuda para el frágil cuidado de la libertad religiosa, sobre todo desde que uno de los aspectos más sensibles en la cuestión religiosa, como son los datos personales, esté siendo altamente protegido por el Derecho nacional e internacionalmente, tal y como refleja el interés de la Unión Europea. Precisamente, es en este espacio europeo donde la exigencia de un ecosistema de confianza ha promovido una regulación más exhaustiva, conforme a las exigencias de respeto a las normas del Estado de Derecho. Sin embargo,

siempre está abierta la posibilidad de que los delitos contra la libertad religiosa puedan aparecer.

Mi propuesta, desde la perspectiva iusfilosófica, parte de que la compleja realidad tecnológica de la IA necesite del principio de precaución, con el que se intenta valorar los riesgos que la IA tiene de cara a la protección de los datos personales de los ciudadanos, especialmente en el ámbito de la protección de la intimidad y de la libertad religiosa, moral o de creencias. En este ámbito, he destacado cómo la IA, lejos de introducir en la actualidad vulneraciones que puedan impedir el desarrollo efectivo de esta libertad, sí puede, en cambio introducir determinados sesgos que la vulneren. El estudio de estos sesgos y su peligrosidad no se ha desarrollado específicamente en el ámbito tecnológico, sino en los delitos de odio que hayan podido expresarse en el entorno digital. Desde este ámbito, los poderes públicos europeos, los más garantistas en la protección de los datos sensibles, han previsto algunos indicadores objetivos que pueden poner de manifiesto los sesgos que identifiquen la vulneración de la libertad religiosa. Así, la OSCE, reconoce la necesidad de la colaboración de las comunidades religiosas para elaborar estos indicadores de sesgos.

Desde esta experiencia previa, considero junto a la Comisión del Parlamento Europeo, el Libro Blanco y las recomendaciones de la propia UNESCO de 22 de noviembre del presente año, que la IA debe desarrollar su actividad en colaboración no sólo con sus programadores y con la Administración pública, sino que, lo que caracteriza el ámbito de la protección de los Derechos y garantías de las libertades en el espacio europeo, se desarrolle también con las entidades jurídico-privadas que permitan un desarrollo sostenible de la confianza de la IA. Para ello, la propia experiencia en los indicadores de sesgos ante el delito de odio, han revelado que, entidades de investigación de las propias comunidades religiosas pueden ser una ayuda quizá imprescindible. Sin embargo, constato cómo un cierto letargo en el desarrollo de estas comunidades, salvo en la religión judía, para adquirir competencias en este ámbito. No obstante, aunque las respuestas desde las mismas entidades religiosas no sean óptimas, la sociedad civil que quiere proteger estos derechos sí parece que se esté movilizando. Sólo si se tiene en cuenta y se capacita formal y materialmente para desarrollar estas competencias, tanto la Administración como los creadores de algoritmos podrán tomar en consideración sus aportaciones, en favor de garantizar el pleno desarrollo de la libertad religiosa.

6. BIBLIOGRAFÍA

Abascal Blanco, Álvaro Juan (2016), *Plataforma de soporte a toma de decisiones frente a situaciones de emergencias en Smart cities*, [Tesis doctoral, Departamento de Ingeniería de la construcción y proyectos de ingeniería, Universidad de Sevilla, Vicerrectorado de Postgrado y Doctorado, Universidad de Sevilla]. https://idus.us.es/handle/11441/39671

Alexandre, Laurent (2018), *La guerra delle intelligenze. Intelligenza artificiale contro intelligenza umana.* EDT.

Aristóteles (1988), *Política,* Gredos.

Asimov, Isaac (1942), *Círculo vicioso*. En https://lecturia.org/cuentos-y-relatos/isaac-asimov-circulo-vicioso/4060/.

Barrio Andrés, Moisés (2020), Tecnologías digitales, ciencia del Derecho y sus desafíos: los principios generales de la Lex Robótica como caso de uso, en: J. Mª. Vázquez García Peñuela - I. Cano Ruíz (eds.), *El Derecho de Libertad religiosa en el entorno digital*, (229-250). Comares, Granada.

Biset, Emmanuel (2017), "Ontología política, esbozo de una pregunta", en: *Nombre* 27, 121-136.

Bubner, Rüdiger (2015), *Polis y Estado. Líneas fundamentales de la filosofía política*, Dykinson, Madrid.

Cicerón (2014), *Los oficios*, Gredos, Madrid.

Cierco Seira, César (2004), "El principio de precaución: reflexiones sobre su contenido y alcance en los derechos comunitario y español", en: *Revista de Administración Pública,* 163 (enero-abril), 73-126.

COMECE (2019), *Robotisation of life*, http://www.comece.eu/comece-publishes-reflection-on-robotisation-of-life [Fecha de consulta: 15 de noviembre 2021]

De Asís Pulido, Miguel (2021), "Derecho al debido proceso e inteligencia artificial", en: Llano Alonso, Fernando, Joaquín Garrido (eds.), *Inteligencia Artificial y Derecho. El jurista ante los retos de la Era Digital*, Thomson Reuters Aranzadi, Navarra, 67-90.

Echániz, R. (2017), *La mezquita como elemento de prevención: el caso de la Ertzaintza*, en: https://www.seguridadinternacional.es/?q=es/content/la-mezquita-como-elemento-de-prevenci%C3%B3n-el-caso-de-la-ertzaintza [Fecha de consulta: 14 de noviembre 2021]

Federación de Comunidades Judías de España, *Informe sobre el Antisemitismo en España durante el año 2019*, en: https://observatorioantisemitismo.fcje.org/ [Fecha de consulta: 14 de noviembre 2021]

Ferreira, Ana Elisabete (2021), "Antropogenia, Principios normativos y Ética artificial", en: Llano Alonso, Fernando, Joaquín Garrido (eds.), *Inteligencia Artificial y Derecho. El jurista ante los retos de la Era Digital*, Thomson Reuters Aranzadi, Navarra, 91-112.

Fioriglio, Gianluigi (2021), "Inteligencia Artificial: retos para un Derecho en la sociedad global", en: Llano Alonso, Fernando, Joaquín Garrido (eds.), *Inteligencia Artificial y Derecho. El jurista ante los retos de la Era Digital*, Thomson Reuters Aranzadi, Navarra, 113-132.

García de Enterría, Eduardo (2005), *Curso de Derecho Administrativo*, Civitas, Madrid.

Guterres, Antonio (2019), *La estrategia y plan de acción de las Naciones Unidas para la lucha contra el discurso de odio*. https://www.un.org/en/genocideprevention/documents/advising-and-mobilizing/Action_plan_on_hate_speech_ES.pdf, [fecha de consulta: 8 de noviembre 2021].

Habermas, Jürgen (2017), *El futuro de la naturaleza humana ¿Hacia una eugenesia liberal?*, Paidós, Barcelona.

Hay, Colin (2006), "Political ontology", en: Goodin, R, Charles Tilly (eds.), *The Oxford handbook of contextual political analysis*, Oxford University Press, Oxford, 78-96.

Hui, Yuk (2020), *Fragmentar el futuro. Ensayos sobre tecnodiversidad*. Caja Negra, Buenos Aires.

Jonas, Hans (2004), *El principio de responsabilidad. Ensayo de una ética para la civilización tecnológica*, Herder, Barcelona.

Latorre Sentís, José Ignacio (2018), *Ética para las máquinas*, Planeta, Barcelona.

Laukyte, Migle (2021), "Dignidad humana y nuevos derechos: El Derecho a la Inteligencia artificial", en: Llano Alonso, Fernando, Joaquín Garrido (eds.), *Inteligencia Artificial y Derecho. El jurista ante los retos de la Era Digital*, Thomson Reuters Aranzadi, Navarra, 183-200.

Llano Alonso, Fernando H. (2018), *Homo Excelsior. Los límites ético-jurídicos del transhumanismo*, Tirant lo Blanch, Valencia.

— (2021), "De máquinas y hombres. Tres cuestiones ético-jurídicas sobre Inteligencia Artificial", en: Llano Alonso, Fernando, Joaquín Garrido (eds.), *Inteligencia Artificial y Derecho. El jurista ante los retos de la Era Digital*, Thomson Reuters Aranzadi, Navarra, 201-234.

Locke, John (1999), "An Magistratus Civilis possit res adiaphoras in divini cultus ritus asciscere, easque populo imponere? Affirmatur", en: Prieto Sanchís, Luis y Betegón Carrillo, *Escritos sobre la Tolerancia,* CEPC, 55-81.

López Oneto, Marcos (2020), *Fundamentos para un Derecho de la Inteligencia Artificial ¿Queremos seguir siendo humanos?* Tirant lo Blanch, Valencia.

Maffetone, Sebastiano (2019), *Politica. Idee per un mondo che cambia*, Le Monnier, Florencia.

Navarro-Valls, Rafael, Javier Martínez Torrón (1997), *Las objeciones de conciencia en el Derecho comparado y español*, Mc Graw-Hill, Madrid.

Nieva Fenoll, Jordi (2018), *Inteligencia artificial y proceso judicial*, Marcial Pons, Barcelona.

Observatorio para la Libertad Religiosa y de Conciencia (2020), *Informe de Ataques a la Libertad Religiosa*, https://libertadreligiosa.es/2021/09/23/aumentan-un-37-los-ataques-a-la-libertad-religiosa-en-espana/ [Fecha de consulta: 9 de noviembre 2021]

OSCE-ODIHR (2009), *Hate Crime Laws, A Practical Guide*. https://www.osce.org/files/f/documents/3/e/36426.pdf [Fecha de consulta: 8 de noviembre 2021].

— (2017), *Desarrollar una comprensión de los delitos de odio de naturaleza antisemita y abordar las necesidades de las comunidades judías,* de 15 de mayo de 2017, https://www.osce.org/files/f/documents/6/d/423680.pdf, [fecha de consulta: 8 de noviembre de 2021].

— (2018), *Delitos de odio contra los musulmanes*, de 22 de febrero de 2018. https://www.osce.org/files/f/documents/6/7/414479.pdf.

Páez Casadiegos, Yidi (2011), *Epifanía y etiología. Ensayos sobre mito y religiosidad griega antigua*, Editorial Universidad del Norte, Barranquilla.

Pérez Luño, Antonio-Enrique (2021), "La inteligencia artificial en tiempo de pandemia", en: Llano Alonso, Fernando, Joaquín Garrido (eds.), *Inteligencia Artificial y Derecho. El jurista ante los retos de la Era Digital*, Thomson Reuters Aranzadi, Navarra, 33-50.

Pinto Fontanillo, José Antonio (2020), *El Derecho ante los retos de la Inteligencia Artificial*. Edisofer, Madrid.

Piñar Mañas, José Luis (2018), *Derecho e innovación tecnológica. Retos de presente y de futuro*, Universidad CEU San Pablo, Madrid.

Van Est, Rinie, Joost Gerritsen, (2017), "Human rights in the robot age. Challenges arising from the use of robotics, artificial intelligence, and virtual an augmented reality", en: *Rathenau Instituut*, Parlamentary Assembly - Council of Europe, Den-Haag.

Tomás de Aquino (1954), *Summa Theologica*, BAC, Madrid.

Turchetti, Mario (2008), "Tiranía y despotismo. Una distinción olvidada", en: Capelli, Guido, Antonio Gómez Ramos (eds.), *Tiranía. Aproximaciones a una figura del poder*. Dykinson, Madrid, 17-58.

Ulloa Rubio, Ignacio (2018), "Libertad religiosa, protección de datos y derecho al olvido", en: *Anuario de Derecho Canónico* 6, 23-58.

Ünver, Halit (2018), *Global Networking, communication and culture: Conflict or Convergence? Spred of ITC, Governance, Superorganism Humanity and Global Culture*, Springer International Publishing, Cham.

Zhao, Houlin (2018), *Carta del secretario general de la Unión Internacional de Telecomunicaciones (UIT) a los participantes en la II edición de la Cumbre Mundial sobre Inteligencia Artificial para el bien*, en Ginebra a 21 de febrero de 2018.

Legislación y Jurisprudencia

Constitución Española (1978) https://www.boe.es/eli/es/c/1978/12/27/(1)/con

Convenio de Oviedo de 4 de abril de 1997, en: https://www.boe.es/eli/es/ai/1997/04/04/(1).

Comisión Europea. (2018). *Inteligencia artificial para Europa* de 25 de abril de 2018, en: https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=COM%3A2018%3A237%3AFIN [Fecha de consulta: 16 de noviembre 2021].

Fiscalía General del Estado, Circular 7/2019 sobre pautas para interpretar los delitos de odio tipificados en el art. 50 CP. https://www.boe.es/diario_boe/txt.php?id=BOE-A-2019-7771, [fecha de consulta: 8 de noviembre 2021].

Parlamento Europeo. (2019). Resolución del Parlamento Europeo de 12 de febrero de 2019 sobre una Política Industrial Global en materia de Inteligencia Artificial y Robótica, https://www.europarl.europa.eu/doceo/document/TA-8-2019-02-12_ES.html#sdocta19 [Fecha de consulta: 16 de noviembre de 2021].

Parlamento Europeo. (2017). Resolución de 16 de febrero de 2017 con recomendaciones a la Comisión en *Civil Law Rules on Robotics* https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html#title1 [Fecha de consulta: 15 de noviembre 2021]:

Real Decreto 932/2013 de 29 de noviembre, por el que se regula la Comisión Asesora de Libertad Religiosa, en: https://www.boe.es/eli/es/rd/2013/11/29/932 [Fecha de consulta: 16 de noviembre de 2021].

Real Decreto 593/2015 de 3 de julio por el que se regula la declaración de notorio arraigo de las confesiones religiosas en España, en: https://www.boe.es/buscar/act.php?id=BOE-A-2015-8642.

Sentencia Tribunal de Justicia Unión Europea. (2014). https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A62012CJ0131.

Tratado de Lisboa por el que se modifican el Tratado de la Unión Europea y el Tratado Constitutivo de la Comunidad Europea de 17 de diciembre de 2007. https://www.europarl.europa.eu/ftu/pdf/es/FTU_1.1.5.pdf

UNESCO, *Proyecto de texto de la recomendación sobre la Ética de la Inteligencia Artificial* de 22 de noviembre de 2021, en: https://unesdoc.unesco.org/ark:/48223/pf0000379920_spa.page=15 [Fecha de consulta 27 noviembre 2021].

III. Robótica e inteligencia artificial jurídica

CAPÍTULO XI

DESAFÍOS IUSFILOSÓFICOS DE LAS ARMAS AUTÓNOMAS[1]

ROGER CAMPIONE
Universidad de Oviedo
campione@uniovi.es

le loup dévore l'agneau, mais il ne le hait pas;
tandis que le loup hait le loup
(Henry de Montherland, *La Guerre civile*)

1. INTRODUCCIÓN

Durante la retirada de las tropas estadounidenses de Afganistán, en el verano de 2021, el Pentágono anunció la neutralización de una 'amenaza inminente' terrorista mediante un *targeted killing*, es decir, un asesinato selectivo realizado con drones *Reaper* que lanzaron un misil *Hellfire* contra un vehículo presuntamente ligado a células del ISIS[2]. El ataque mató a diez personas. En realidad, el objetivo militar resultó ser un Toyota Corolla conducido por un trabajador local de una ONG americana, Zemari Ahmadi, que acababa de llegar a casa tras haber repartido alimentos en campamentos de desplazados. Junto con el conductor 'sospechoso', las víctimas fueron todos miembros de su familia, de los cuales siete niños. Las autoridades políticas y militares estadounidenses, como había ocurrido en otras ocasiones, no encontraron motivos para sancionar a los responsables implicados en el ataque, reconduciendo lo acontecido a la conocida categoría bélica de los daños colaterales. Las explicaciones proporcionadas por los mandos a la prensa, para evidenciar la ausencia de eventuales incumplimientos normativos, se basaban en que la decisión de lanzar el ataque había sido tomada atendiendo a un programa de inteligencia artificial para el cual existía una 'certeza razonable' de que ese era el objetivo correcto. Y la respuesta parecería implicar el argumento de cómo se le va a exigir que rinda cuenta de sus decisiones a un algoritmo. ¿Sería eso posible en términos de responsabilidad jurídica?

El proyecto *Maven* es un sistema basado en un algoritmo creado para las fuerzas armadas estadounidenses y con el objetivo de alertar a los militares

[1] Este trabajo se enmarca en el Proyecto Prueba de concepto SMARTWAR. Viejas guerras y nuevas tecnologías: un banco de prueba para la regulación de la violencia política (PDC2021-121472-I00).

[2] Sobre la teoría, la práctica y la ausencia de previsión normativa en el derecho internacional de los asesinatos selectivos en general y mediante el uso de drones en particular, remito para una síntesis a Aldave, 2017, 171 ss.

cuando aparecen ciertos objetos en un territorio sometido a su vigilancia. Dado que a los operadores humanos les resultaría imposible ver y catalogar todas las imágenes y las grabaciones recogidas en un determinado teatro de operaciones militares, se emplea ese algoritmo que tiene la capacidad de analizar esa tan ingente cantidad de datos gráficos y avisar cada vez que detecta, por ejemplo, un vehículo o cierto individuo. Y su ámbito de aplicación puede extenderse al control de otro tipo de datos, ampliando sobremanera las posibilidades de organización, planificación y logística de las operaciones bélicas por parte de las fuerzas armadas (Noël, 2020, 66).

Al margen de que la historia de Zemari Ahmadi demuestre que detrás de la supuesta evolución tecnológica se esconden los intentos más viejos y tradicionales de escurrir el bulto por parte de humanísimos sujetos cuando toman decisiones que no saben o no pueden justificar, no es menos cierto que conviene prestar atención a las implicaciones normativas de la inteligencia artificial, al menos por dos razones.

La primera es que los avances en robótica e inteligencia artificial ya se están adelantando al uso de drones, incluyendo sistemas autónomos cada vez menos dependientes de la intervención humana a la hora de tomar decisiones. Como manifestó hace años Philip Alston, en un previsible futuro existirá la tecnología para crear robots capaces de seleccionar y actuar con una mínima intervención humana o incluso sin necesidad de ella (United Nations, 2010, 10). Además, hay razones económicas que empujan a ello, pues los sistemas autónomos resultan más ventajosos que los drones controlados en remoto, cuya utilización es, a su vez, más barata que la de los aviones tripulados.

La segunda razón por la que merece atención el tema tiene que ver con la propia genealogía de la experiencia jurídica: *hominum causa omne ius constitutum est*, decía Hermogeniano entre finales del siglo III y comienzos del IV: todo el derecho se ha creado por causa del hombre. Pero el desafío lanzado por las tecnologías convergentes puede acabar deshilachando definitivamente el cordón umbilical que ata el derecho al sustrato humano del sujeto y que, desde las *Instituciones* de Gayo, prioriza sistemáticamente a la persona frente a las cosas y las acciones. O, al menos, su impacto obliga a reavivar en profundidad la reflexión acerca de lo que significa *persona* en derecho[3].

[3] El tema, obviamente, es de tal amplitud y calado como para ocupar no solo una obra autónoma, sino toda una línea de investigación filosófico-jurídica. Para un simple y muy somero arranque en la materia me permito remitir a Campione, 2005, 184 ss., también para alguna indicación bibliográfica de partida.

2. DERECHO Y GUERRA

Partamos de una antropología mínima: el derecho es antropomorfo en el sentido de que, amén de las diferencias entre culturas y enfoques disciplinares, solo cabe hablar de derecho si implicamos en ello como algo normal una percepción, por parte de los sujetos, de lo que el derecho dispone en el caso concreto, ordenando, permitiendo o prohibiendo (Sacco, 2007, 19)[4]. Percibiendo así que el orden artificial del derecho se sostiene sobre los individuos que lo producen, aunque estos se perciban como sujetos solo porque una norma los reconoce como tales (Barcellona, 1996, 47). Y apreciando las desviaciones de la regla, en consecuencia, como una quiebra de la armonía social. Un derecho, por tanto, siempre presente, pero, en toda la esfera de las acciones lícitas, un callado compañero de viaje (Corradini, 1995, 17).

Hay otras opciones, naturalmente. Escojo esta, quizá, porque me permite plantear con más eficacia algunas cuestiones que afectan a la relación del ser humano y del derecho con la técnica, lo cual representa el contexto general en el que pretendo moverme. En particular, es así porque en estas páginas me referiré principalmente a problemas que afectan al derecho desde una perspectiva internacional, en cuyo tablero es menos tentador recurrir a la vinculación entre derecho y Estado -siempre que fuera posible dar de este una definición unívoca- o invocar la presencia de una autoridad superior real y claramente identificada. Y donde es más complejo disponer de alguna categoría normativa y/o conceptual para afrontar en la práctica la resolución de los conflictos o la imposición de sanciones.

No es que quiera aventurarme en la *inveterata quaestio* teórica de si es realmente derecho el derecho internacional (Hart, 2012, 214; Ferrajoli, 2007, 482-483). Más simplemente, parto de la amplia premisa zoliana según la cual "en la arena internacional la relación entre derecho y poder es tan estrecha y tan ambigua que una filosofía del derecho internacional se vería reducida a una simple especulación normativa si no colocase en el centro de su teoría las muchas variables que tornan problemática la relación entre el derecho *in books* y el derecho *in action*; es decir, si no estudiase como objeto específico de la 'ciencia jurídica' la red de transacciones políticas, económicas y sociales mediante las cuales los principios y las reglas del derecho se convierten en disciplina efectiva de casos concretos" (Zolo, 1998, 138). En consecuencia, habría que aceptar la imposibilidad de una teoría pura del derecho internacional con más rotundidad, si cabe, que respecto del derecho estatal. Por otro lado, esta

[4] Dejo abierta la cuestión, tanto jurídica como antropológica, relativa a la posibilidad de hablar de derecho en relación con ciertas sociedades animales evolucionadas. Como escribe el mismo Sacco, se trata de una elección subjetiva (*ibid.*).

premisa no niega la relevancia fáctica de las normas internacionales. Como he defendido en otro lugar, que se comparta la idea kelseniana del derecho internacional como ordenamiento jurídico estructuralmente primitivo, se coincida con Ross en que su imperfección es accidental o se esté de acuerdo con Hart en que el derecho internacional no conforma un auténtico sistema jurídico, en ningún caso cabe colegir que tal derecho sea necesariamente inane desde el punto de vista práctico (Campione, 2020, 81-82)[5].

El banco de prueba del derecho internacional se torna especialmente exigente en lo relacionado con el uso de la fuerza en los conflictos armados. En contra de lo que podemos asumir como un lugar común, no es cierto que la guerra haya existido siempre como forma de aniquilación del enemigo. Al contrario, este es un fenómeno más bien contemporáneo producido por los potentísimos medios destructivos creados por la tecnología militar. Las guerras tradicionales, hasta comienzos del siglo XX, consistían en enfrentamientos circunscritos de ejércitos profesionales. A diferencia de las guerras clásicas, sometidas a los límites objetivos de los medios militares, la guerra contemporánea, tanto convencional como nuclear, anula esas fronteras desbaratando la lógica de la guerra circunscrita a ciertos sujetos y espacios. Y, por supuesto, a ciertas normas[6]. Así y todo, desde el punto de vista normativo, hasta comienzos del siglo XX la guerra era un 'derecho natural' que todo Estado tenía en el marco del *ius publicum europaeum*. El derecho de guerra cubre históricamente dos ámbitos, el del por qué se puede hacer (*ius ad bellum*) y el del cómo debe hacerse (*ius in bello*). En la época antigua y medieval la cuestión se

[5] En esa sede me refería especialmente al *ius in bello* o derecho internacional humanitario. Entre los varios lugares en que Kelsen expresa esa idea acerca del ordenamiento internacional, cfr. Kelsen, 1952, 36: "The obvious insufficiency of reprisals and war as sanctions of International law is the consequence of the complete decentralization of the community constituted by this law, which, precisely because of this decentralization, and especially because of the lack of a centralized executive power, has the typical character of a primitive law"; de acuerdo con Ross, 1947, 19: "as long as we are concerned with an international order, that is to say, an order directly concerned with states, not individuals, it will, however, be necessary to draw a radical distinction between International Law and Internal Law. The present lamentable conditions in this domain are not implicit in the nature of International Law. International Law is not «conceptually» but only «accidentally» imperfect law"; y en el caso de Hart, 2012, 236: "it is submitted that there is no basic rules providing general criteria of validity for the rules of international law, and that the rules which are in fact operative constitute not a system but a set, among which are the rules providing for the binding force of treaties".

[6] Según Ferrajoli, si "esta es hoy la guerra, el paradigma de la guerra como sanción o reparación es del todo inutilizable". En primer lugar, porque se convierte en un castigo infligido a inocentes -las poblaciones civiles- y, en segundo lugar, porque su desmesura siempre acaba provocando reacciones desproporcionadas ante cualquier violación del ordenamiento internacional (Ferrajoli, 2004, 60). En la misma línea, con una mayor *vis* crítica, Zolo, 1998. Para Bobbio, la devastación de la guerra termonuclear incluso mermaba la posibilidad de justificar una guerra en legítima defensa (Bobbio, 1992, 55-56).

centraba esencialmente en el *ius ad bellum*, el cual dependía de la doctrina de la 'guerra justa', basada en la existencia de una justa causa, es decir, de un requisito moral y sustantivo: haber sufrido una injusticia.

Con el moderno sistema de Estados nacionales, soberanos, surgido de la paz de Westfalia, se consolida la idea de que ese requisito de la justa causa presenta un problema irresoluble: ¿cuál de los dos bandos tiene de su lado la justa causa? ¿Quién sufre la injusticia? ¿Quién decide si los Estados (*que superiorem non recognoscens*) tienen razón o no? Europa ya no es la República cristiana, ya no basta con decir *Deus vult*, como hizo Urbano II en Clermont, para poner a todos de acuerdo. La idea de que en cualquier guerra cada bando enfrentado entiende que tiene la razón de su parte es tan obvia que a menudo se olvida. Basta con echar un vistazo a la actualidad internacional para comprobar que incluso en conflictos militares donde resultan estar claros los papeles de quien ha asestado el primer golpe, como Rusia contra Ucrania, no faltan repertorios de argumentos públicos para justificar lo que se está haciendo no solo por parte de Ucrania y sus aliados, sino también por parte de Rusia, que expone sus propios agravios. Es algo que ha pasado, está pasando y pasará en todas las guerras.

Con la consolidación del *ius publicum europaeum*. se abandona entonces la idea de la 'justa causa' y la guerra pasa a ser un atributo de la soberanía, es decir, no puede ser calificada como justa ni injusta. La decisión de declarar la guerra es una libre disposición del Estado. Ese derecho público europeo había dejado atrás la doctrina medieval de la guerra justa y lo que imponía era el cumplimiento del *ius in bello*, o sea, de las reglas de conducta de la guerra respecto del enemigo. La atención hasta lo que se llamará derecho internacional humanitario se explica aún mejor si se tiene en cuenta el importante desarrollo de la tecnología militar de la época. Se trata de una nueva vertebración normativa que se inserta en un modelo clásico de guerra, transmitido por la épica antigua y la fascinación ancestral del duelo, que 'encorsetaba' -en el sentido de limitar las acciones- la contienda dentro de las reglas impuestas por la ética caballeresca.

Este desarrollo tecnológico provoca desajustes entre las incipientes tácticas bélicas de la modernidad y los códigos de honor militar tradicionales. Pensemos en eventos bélicos como el que aconteció durante la batalla de Pavía en 1525, cuando el rey de Francia, Francisco I, fue capturado mientras lideraba sus tropas, junto a la flor de su nobleza, y perdió la batalla por haberse situado a la cabeza de sus arcabuceros obstruyendo, por tanto, la línea de tiro de sus propios trabucos que no podían apuntar al enemigo para no derribar a los propios. El rey francés actuó siguiendo esquemas bélicos puramente medievales, pretendiendo ganar la batalla con honor y gloria. Sin embargo,

Pavía marca un punto de inflexión en la relación estratégica entre infantería, caballería y artillería. Durante la Edad Media la caballería pesada había constituido la columna vertebral de los ejércitos, pero ya antes del siglo XVI esta disposición empezó a cambiar y precisamente en las guerras de Italia culminó la evolución del arte bélico renacentista que no afectó solo a la caballería, sino también a las nuevas estrategias utilizadas por las unidades de infantería llamadas a hacer frente a las amenazas de las nuevas piezas de artillería. El conocimiento de las armas es un dato imprescindible para comprender la guerra y las relaciones entre sus actores.

3. BREVÍSIMO PARÉNTESIS SOBRE LA ACTUALIDAD

El concepto de enemigo es fundamental desde el punto de vista jurídico: como escribe Sánchez Ferlosio, el núcleo germinal del derecho internacional hay que buscarlo en los deberes para con el enemigo (Sánchez Ferlosio, 2016, 242). El enemigo es un espejo jurídico: la primera y fundamental regla de la guerra es que el enemigo tiene nuestros mismo derechos y deberes. En el contexto del derecho público europeo este es el marco normativo dirigido a evitar los efectos más destructivos de la guerra, centrándose en limitar qué se puede y no se puede hacer, al margen de quien se considere que tiene más razón moral, política, religiosa, ideológica, etc.[7]

Una vez que la experiencia de las guerras mundiales conduce a declarar la guerra un crimen en el siglo XX, esa hermenéutica *polemológica* entra en crisis. Esto ocurre de modo especial a finales del siglo XX, debido a la recuperación de modelos de legitimación de la guerra basados no el respeto de las reglas de conducta sino en la justa causa material. Dos experiencias en particular alimentan ese *revival*: la culminación del concepto de 'guerra humanitaria', consolidado por la OTAN en los Balcanes, y la Guerra global contra el terrorismo desencadenada a raíz del 11-S. Estos tránsitos han apuntalado el regreso de esquemas argumentativos mutuados de la doctrina de la guerra justa, donde no se enfrentan dos beligerantes en igualdad de condiciones jurídicas sino dos bandos que representan el bien y el mal, con la consecuencia de que lo hecho por unos durante la guerra está bien y lo que hacen los otros está mal, aunque sea lo mismo. Vuelven a primar los valores morales de quien emite los juicios frente a las reglas jurídicas de conducta válidas en la misma medida para todos.

Algunos, como escribía a propósito del conflicto en Ucrania Zagrebelsky, creemos que los poderosos que en tiempo de guerra enarbolan una superioridad moral no favorecen la paz. Estamos ante una guerra (que, por

[7] Sobre el concepto de enemigo y su transformación en los nuevos escenarios bélicos, véase la reciente contribución de Rodríguez Fouz, 2022.

cierto, no es la única). Estamos ante algunas guerras: están los agresores y están los agredidos. Esta es la única certeza sobre la cual no hay dudas acerca de la invasión rusa. Hay víctimas civiles que han de ser ayudadas, pero también hay que contrastar, sostiene Zagrebelsky, las ideas agresivas que prefiguran un futuro quizá peor que el presente y que, en cualquier caso, alejan la perspectiva de un posible entendimiento que ponga fin a la guerra. Sobriedad y espíritu crítico, por tanto, no para negar la evidencia sino para evitar lo peor (Zagrebelsky, 2022).

4. EL OFICIO DE LAS ARMAS

La formación del Estado moderno, entre el siglo XVI y XVII, se vio notablemente impulsada por las guerras libradas para dominar Europa. Los ejércitos y las grandes flotas empleadas en los cruentos conflictos armados de la época implicaban unos costes bélicos muy altos, tanto desde un punto de vista económico como humano. Sus crecientes exigencias financieras solo se podían gestionar consolidando una estructura administrativa, burocrática y fiscal cada vez más centralizada, que desembocó en la forma Estado. En esta nueva estructura los ejércitos permanentes garantizan la fuerza política del rey no solo durante las contiendas bélicas. Tanto si se trata de tropas aposentadas que guarnecen castillos o pasos fronterizos, como si hablamos de batallones móviles listos para desplazarse en cualquier momento, estas fuerzas militares cumplen su función de protección real también en tiempos de paz. A partir de la segunda mitad del siglo XV, en Italia, las nuevas formas del arte de la guerra, ligadas a las transformaciones técnicas de las milicias (básicamente mercenarias y dependientes únicamente de las arcas reales), inciden profundamente en la vida y el desarrollo de los Estados. De este modo, de cara a la política exterior, el monarca dispone de recursos que sería impensable alcanzar de otra forma y, a la vez, con esta autonomía económica se desvincula de la presión política feudal. "Las nuevas formas del arte de la guerra que se imponen a partir de la segunda mitad del siglo XV", recuerda Federico Chabod, "constituyen el aspecto técnico de una profunda alteración que, a través de la técnica, incide hondamente en la vida del Estado" (Chabod, 1967, 603)[8].

[8] Es rotunda la crítica de Maquiavelo al empleo de las milicias mercenarias y prestadas por terceros: son inútiles y peligrosas y son la causa de la ruina de la Italia de su tiempo que, por muchos años, confió en las armas mercenarias. Es conocido el pasaje del capítulo XII de *El Príncipe* en el que el Secretario florentino afirma que disponer de buenas armas era un principio fundamental de mantenimiento de la estabilidad política, pues son necesarias, para ello, buenas leyes y buenas armas. Mas debido a que no puede haber buenas leyes sin buenas armas, es de estas que conviene hablar. Y el problema de Italia fue debido a que, habiendo acabado en manos de la Iglesia y de alguna república, por tanto, de curas y ciudadanos no acostumbrados a las armas, empezaron a recurrir a forasteros. Un príncipe sabio, dice Maquiavelo, rehúye de estas

(...)

Si bien en el primer cuarto del siglo XVI seguía teniendo relevancia táctica la infantería ligera de tradición itálica y española, las armas de fuego manuales fueron las que redibujaron por completo el marco de empleo de las fuerzas militares de infantería.

Lo cuenta magistralmente Ermanno Olmi en *Il mestiere delle armi*: en 1526 una armada imperial de lansquenetes luteranos baja por el norte de Italia con la intención de saquear Roma para castigar al Papa Clemente VII, reo de haberse aliado con Francisco I de Francia en la Liga de Cognac contra el Sacro Romano Imperio. Lidera la vanguardia de las tropas pontificias Juan de Médicis (*Giovanni delle Bande Nere*) quien, disponiendo de menos hombres, adopta la táctica militar de las refriegas para debilitar las cohortes imperiales. La noche del 25 de noviembre *Giovanni delle Bande Nere* ataca a los soldados del general Frundsberg, pero se encuentra con una inesperada sorpresa: los germánicos esconden, tras las barricadas de ladrillo, unos falconetes (nuevos cañones de pequeño calibre) que alcanzan al afamado *condottiero* y cuya herida le causará la muerte a los pocos días. El oficio de las armas, que le había aupado a la gloria, se ve superado por la irrupción del nuevo invento, pues hasta la armadura se convierte en un ropaje inútil frente a la potencia de fuego del explosivo cañón[9]. Un desarrollo tecnológico que, al modificar las formas del arte militar, provoca cambios definitivos también en el armazón normativo y axiológico de la guerra y marca el ocaso de toda una ética caballeresca centrada en los valores individuales del duelo bélico o la habilidad táctica del estratega. La posibilidad de golpear desde lejos, de evitar la lucha cuerpo a cuerpo, hace depender la victoria de la capacidad material (y, por tanto, económica) de fuego y sella la irrupción de un marco militar 'robótico', en el sentido de que el impacto de la fuerza bélica se traslada del ser humano al arma, a la máquina. Tradicionalmente, los combates tenían reglas que constituían un código de honor respectado por todas las partes. La introducción de estas armas nuevas provoca una progresiva despersonalización del oficio de la guerra, otorgando un poder cada vez mayor a los *artefactos* de la inteligencia, es decir, a los productos de una inteligencia *artificial*.

Las consecuencias de estas mutaciones, desde un punto de vista histórico, político, jurídico e incluso antropológico, no pueden ser ignoradas. Por un lado,

armas y confía solo en las propias, "iudicando no vera vittoria quella che con le armi aliene si acquistasse" (Machiavelli, 1949, 38-44).

[9] En realidad, si nos ceñimos a las fuentes, parece ser que fue un mosquete, es decir, un arma de manejo manual, la que destrozó la pierna del héroe, y no un pequeño cañón. Sin embargo, la historiografía del siglo XIX, más proclive a la narrativa que a la mera crónica, no pudiendo engrandecer las gestas de Giovanni delle Bande Nere, se habría sentido obligada a aumentar notablemente las dimensiones de la boca de fuego que había truncado la gloriosa carrera del *condottiero* (Scalini, 2001, 122).

a partir del siglo XVI, la introducción de las armas de fuego pondría en marcha un proceso imparable de transformación de la estrategia militar pues, como recuerda Pietro Aretino en la película, "las nuevas armas de fuego cambian la guerra, pero son las guerras las que cambian el mundo". Por el otro lado, en aquella época se va afirmando y consolidando un nuevo modelo de conducta normativa para el soldado, pero también para el político. De hecho, junto con el aspecto militar, durante el Renacimiento el cambio se nota especialmente gracias al advenimiento de una diplomacia profesional que anuncia una nueva época en la política exterior, pareja al surgimiento del llamado principio de equilibrio de potencias europeas. Al fin y al cabo, también en este caso se trata de una innovación técnica, debida a un cambio institucional de estructura y función, como es la residencia estable y permanente de representantes de un país en el territorio de otro país, al margen de cuál sea la relación contingente entre los Estados. Claro que la novedad 'técnica' acarrea un progreso 'político', representado por la toma de conciencia de la importancia alcanzada por las relaciones internacionales.

La aparición de la artillería, en el siglo XVII, junto con la renovación de la infantería, llevó a la creación de instituciones especializadas en la producción de armas y en la organización de campañas militares más prolongadas. En este contexto de transformación tecnológica, solo los Estados tienen capacidad para acometer y sostener en el tiempo los gastos exigidos por el desarrollo de las actividades bélicas (Noël, 2020, 67). No sorprenden, por tanto, afirmaciones como la de Tilly: "la guerra hace al Estado y el Estado hace la guerra" (Tilly, 1975, 42). Ni que se nos explique la existencia de un profundo nexo semántico entre los vocablos más antiguos que los griegos usaban para indicar respectivamente la agregación política y el conflicto armado, *polis* y *polemos* que, en su forma arcaica *ptolis* y *ptolemos* compartirían la raíz *pt* (Miglio, 2016, 26)[10].

Sin embargo, como decía al principio, la evolución tecnológica de la guerra habría conducido a un punto en el que esa conexión etimológica y política colapsa, provocando un cortocircuito que volvería imposible la convivencia histórica entre sociedad y guerra. Es archiconocida la imagen de Bobbio, para quien "la guerra moderna se ubica fuera de todo posible criterio de legitimación y legalización, más allá de todo principio de legitimidad o de

[10] Precisamente sobre la base de esta premisa etimológica, no comparto la posterior definición que Miglio da de la guerra como 'ausencia de límites para destruir o someter al enemigo' (Miglio, 2016, 38 ss.). Así como la *polis* remite a una forma de agregación necesariamente 'política', la guerra (*polemos*) solo puede ser identificada como tal si se asume que posee una determinada 'forma'. No toda destrucción del enemigo es guerra, sino solo la que se libra en el marco de las reglas de la guerra. Lo contrario es la aniquilación del enemigo sin encomiendas, sin rituales, sin avisos, sin generales, sin soldados, sin civiles, sin beligerantes establecidos; es violencia indistinta, *nuda*, difusa, sin límites, sin control.

legalidad. Es incontrolada e incontrolable por el derecho, como un terremoto o una tormenta. Después de haber sido considerada bien como un medio para realizar el derecho (teoría de la guerra justa) bien como objeto de reglamentación jurídica (en la evolución del *ius belli*), la guerra vuelve a ser, como en la representación hobbesiana del estado de naturaleza, la antítesis del derecho" (Bobbio, 1992, 60). Hoy más que nunca parece estar nuevamente de actualidad el temor ante el horror apocalíptico de una potencial guerra nuclear; sin embargo, al mismo tiempo, en las semanas en que han sido redactadas estas líneas, asistimos a una guerra que recuerda las del siglo XX, pero librada con la tecnología del siglo XXI. En efecto, es lo que viene a ser una guerra híbrida, de última generación. Y la utilización de estas tecnologías está planteando desafíos -también normativos- que no se identifican de manera específica con el panorama aterrador prefigurado por el hongo atómico.

5. DE AUTOMÁTICO A AUTÓNOMO

De hecho, en los últimos tiempos, los drones, los sistemas de armamento autónomo, (*Lethal Autonomous Weapon Systems* - LAWS), las formas de potenciamiento humano mediante la técnica, los sistemas de inteligencia artificial han puesto sobre la mesa de quienes reflexionan sobre las formas de la guerra cuestiones que no sabemos muy bien si y hasta qué punto definir como nuevas. Porque con ellas no valen las mismas razones no apocalípticas que Bobbio aducía para destacar las líneas de ruptura de la guerra nuclear con la guerra del pasado (Bobbio, 1992, 32-36). A diferencia de aquella, la guerra protagonizada por sistemas militares dotados de diferentes grados de autonomía no copa la atención de los especialistas porque ponga en peligro a la humanidad entera, sino porque no aclara los márgenes de actuación bélica de tales máquinas, volviendo opaco el funcionamiento de los procesos decisionales, el nexo de imputación de actos y, por tanto, la cadena de responsabilidad técnica, política y militar. A diferencia de la guerra nuclear, el carácter no inmediatamente aniquilador del empleo de sistemas autónomos no vuelve inoperantes las teorías y doctrinas elaboradas para justificar la guerra. Y finalmente, desde un punto de vista utilitarista, la guerra librada con armas de inteligencia artificial no implica necesariamente la imposibilidad de distinguir al vencedor de la contienda, tal como ocurriría con el nivel de devastación causado por un conflicto atómico.

La automatización militar ha ido cambiando de forma progresiva, especialmente en los últimos dos siglos, el modo de guerrear y de afrontar los múltiples y generalmente insolubles problemas éticos, políticos, jurídicos y económicos que los conflictos armados plantean. En los últimos años, los avances registrados en el proceso de automatización nos han proyectado, casi

de sopetón, hacia un *topos* ulterior, el del aprendizaje autónomo y la inteligencia artificial.

Aquí ya no se trata de medir simplemente el creciente impacto de técnicas e instrumentos que amplifican la capacidad de impacto militar de ejércitos o soldados, sino de indagar el horizonte más realista entre dos polos: por un lado, la progresiva desaparición del soldado humano en el campo de batalla, según las visiones más 'futuristas', reemplazado por una máquina no expuesta a los factores imponderables que caracterizan el proceso humano de toma de decisiones; por el otro, la necesidad de rechazar de plano el empleo de los sistemas autónomos en las contiendas para evitar que la guerra y sus muertes dependan de algoritmos inhumanos en lugar que de seres vivos.

El arte de la guerra está cambiando de nuevo también porque la evolución tecnológica ha impactado de forma abrumadora en la información, es decir, en la narrativa de la guerra. Estamos observando la abundancia de información sobre el conflicto de Ucrania y lo costoso que es diferenciar la verdadera de la falsa, si es que tiene todavía algún sentido esa distinción en la era de los ciento cuarenta caracteres, el *TikTok* y, sobre todo, la posibilidad de construir relatos audiovisuales ficticios imposibles de detectar, gracias a los avances de las técnicas digitales.

La aplicación y expansión de la inteligencia artificial en la dimensión militar podrían alimentar el comienzo de una transformación semejable a la experimentada por el Estado en el siglo XVI. En aquel entonces, la percepción del cambio respecto de los estándares de comportamiento en los conflictos armados, por seguir citando la película de Ermanno Olmi, se reflejaba en la crisis 'climática' acarreada por la nueva geografía bélica: la verdadera desgracia, lamentan los caballeros de la incipiente modernidad que protagonizan la película, es que ya no se da tregua a la guerra, ni tan siquiera en invierno, cuando el frío y el mal tiempo golpean más fuerte que las armas. En la última Resolución sobre "Inteligencia artificial en la era digital", el Parlamento Europeo destaca la necesidad de una nueva orientación en esa crisis climática, considerando "que los sistemas armamentísticos basados en IA deben estar sometidos a normas globales y a un código ético de conducta internacional para respaldar el despliegue de tecnologías de IA en operaciones militares, dentro del pleno respeto del Derecho internacional humanitario y el Derecho en materia de derechos humanos, y en consonancia con el Derecho y los valores de la Unión" (Parlamento Europeo, 2022, 20)[11]. Volviendo al paralelismo con el Renacimiento, se trata de una fórmula más genérica y menos decidida de la

[11] La Resolución, de 3 de mayo de 2022, está disponible en https://www.europarl.europa.eu/doceo/document/TA-9-2022-0140_ES.pdf

utilizada por un palafrenero de las milicias pontificias, tras consumarse el saco de Roma por los Lansquenetes en mayo de 1527: "por motivo de la siniestra suerte tocada al Señor Juan De Médicis, los más ilustres capitanes y comandantes de todos los ejércitos formularon buenos auspicios para que nunca jamás se volviera a utilizar contra el ser humano la poderosa arma de fuego". Fue un presagio, desde luego, poco realista. Sin embargo, no falta quien sostiene que la misma existencia de las armas nucleares tuvo el efecto disuasorio de que todos quisieran evitar la guerra, de modo que podríamos preguntarnos si el imparable proceso de sofisticación de la tecnología militar produciría armas inteligentes de tal calibre destructivo que desalentarían definitivamente cualquier atisbo de guerra (Tegman, 2018, 141).

Lo que me parece acertado, como decía Bobbio en su autobiografía, es que la historia no terminó tal como profetizó Fukuyama; posiblemente, muy al contrario, si centramos nuestra atención en el progreso tecno-científico que ha transformado -está transformando, escribía todavía Bobbio a finales de los Noventa- la comunicación y, por tanto, las relaciones humanas, nos damos cuenta de que la historia acaba de empezar. Lo verdaderamente difícil es vislumbrar qué dirección ha enfilado, ya que la creación de nuevos instrumentos cada vez más eficientes abocan a una 'revolución permanente', en el sentido de una transformación radical que no deja lugar para la vuelta a ningún *statu quo* anterior. Que esto represente algo bueno o malo, no nos es dado preverlo pues, como recordaba el mismo Bobbio en las páginas finales, la ciencia del bien y del mal aún no ha sido inventada. Así que la influencia del progreso tecnológico en la calidad de la vida pública y de la democracia es imprevisible por ambivalente. Más que ayer, hoy podemos rubricarlo con sus últimas palabras: "ni siquiera sabemos si somos dueños de nuestro destino" (Bobbio, 1997, 257-261).

La incertidumbre está acrecentada por el hecho de que no existe necesariamente una proporcionalidad entre el grado de innovación tecnológica y la envergadura de los efectos que puede tener. En el ámbito militar, una pequeña evolución técnica como la sustitución del brazo mecánico de un robot utilizado para bonificar campos de minas por una torrecilla ametralladora, plantea una gran cuestión ético-jurídica. Las fuerzas armadas estadounidenses usaron en Irak la primera versión de un robot llamado *Talon* en 2004 para desminar y a partir de 2007 emplearon también la versión de combate llamada *Talon Sword*. Si bien armas de este tipo no pueden ser definidas autónomas porque les faltan dos condiciones, no pueden seleccionar ni atacar de manera autónoma un objetivo militar, es comprensible que desde hace unos años se hayan activado los 'centinelas éticos' ante la perspectiva de un vacío de responsabilidad provocado por la desaparición de un *alguien* al que pedir

cuenta por las consecuencias ilegales o inaceptables de una decisión tomada por un *algo* (Tamburrini, 2020, 77).

El debate está sobre la mesa porque los *killer robots*, es decir, los sistemas de armamento autónomo no son ya ciencia ficción. Drones como el *Phantom 4* o el Mavic Pro, fabricados por la compañía china DJI, por ejemplo, poseen una tecnología que les permite, de forma autónoma, reconocer objetivos y evitar obstáculos; por tanto, pueden seguir a un vehículo sin necesidad de comunicarse con el controlador humano remoto. Dispositivos como el estadounidense enjambre de drones *Perdix* no son aparatos pre-programados individualmente, sino 'organismos colectivos' que comparten un cerebro repartido para tomar decisiones y adaptar recíprocamente su conducta unos en función de otros, como los enjambres en la naturaleza. La prensa británica ha hablado del *Taranis*, un avión de combate desprovisto de piloto, invisible para los radares con incluso la posibilidad de atacar de forma autónoma[12].

6. LOS SISTEMAS AUTÓNOMOS A DEBATE

Los avances técnicos logrados en el campo de la inteligencia artificial apuntan a un posible y paulatino remplazo de los soldados en el campo de batalla. Ante esta perspectiva hay quienes opinan que ese desarrollo técnico debería ser normativamente vedado, para evitar que conduzca a una autonomía plena del sistema de armamento autónomo. Una prohibición preventiva sería factible y sería la única actitud posible para afrontar con éxito el peligro potencial representado por los LAWS. Otros, sin embargo, argumentan que estas 'máquinas' podrían ser intérpretes más fiables y rigurosos de los estándares éticos y jurídicos preestablecidos. Que el diseño y el uso de los sistemas autónomos podría llevar potencialmente a salvar un mayor número de vidas de no combatientes. Varias son las razones aducidas por ambas posturas.

En el primer sentido, se sostiene que el empleo de sistemas armamentísticos que en su funcionamiento no involucran a sujetos humanos no constituiría simplemente un nuevo tipo de arma, sino un nuevo método de guerra que modificaría radicalmente, y con peores resultados, las formas bélicas conocidas. También habría que tener en cuenta el clásico 'dilema de la seguridad', conforme al cual, si un país se hace con armas de este tipo, otros países se sentirán impelidos a imitarlo, desencadenando una escalada militar robótica. Además, existiría la posibilidad concreta de que los *killer robots* fueran

[12] Previendo la eventualidad de que cambien las *rules of engagement* (reglas de enfrentamiento) que actualmente siguen exigiendo la autorización de un operador humano en todo uso de la fuerza (https://www.dailymail.co.uk/sciencetech/article-3634980/RAF-drones-kill-without-need-human-operators-AI-let-machines-pick-targets-fire-will.html; <https://www.thetimes.co.uk/article/raf-drone-could-strike-without-human-sanction-mzpjmr786>; (Tamburrini, 2020, 83).

adquiridos por regímenes represivos y no especialmente respetuosos con las leyes internacionales. Sin olvidar que con los LAWS aumentaría la eventualidad de ataques armados, puesto que disminuiría el riesgo de pérdidas humanas en combate y esto tendría efectos nocivos sobre la seguridad internacional. Los problemas normativos también serían patentes: ¿cómo podrían las máquinas cumplir con disposiciones como la Cláusula Martens[13], interpretando y recurriendo a los usos de las naciones civilizadas, los principios de la humanidad y las exigencias de la conciencia pública? Habría serias dudas de que pudieran cumplir con el derecho de los conflictos armados. Y qué decir de los serios problemas de responsabilidad cuando sea un sistema autónomo el que incumple esas normas. A quién imputársela: ¿al operador, al oficial al mando, al programados, al fabricante? Por todo ello, la mejor forma de estigmatizar el uso de los sistemas de armamento autónomos sería un tratado internacional específico que estableciera una prohibición absoluta (Goose, 2015, 43-45)[14].

Desde el otro lado, la perspectiva adoptada parte de una triste y realista premisa: que el fenómeno de la guerra seguirá existiendo y que lo principal será proteger a los no combatientes de una forma mejor que hasta ahora. El *statu quo* con respecto a los civiles sería en la actualidad inaceptable y habría que implementar instrumentos de mejora. En este sentido pesimista, más fácil que reformar las tendencias milenarias del comportamiento humano será fabricar robots que superen las expectativas de conducta moral de los humanos en el campo de batalla. Sería creíble, por tanto, que las máquinas se mantuvieran más pegadas que las personas a los requerimientos del derecho internacional humanitario[15]. Y, en definitiva, los sistemas autónomos podrían actuar de una manera más ética que los soldados de carne y hueso. Por ello, frente a la

[13] La cláusula Martens forma parte del derecho de los conflictos armados desde que apareciera en el Preámbulo del II Convenio de La Haya de 1899 relativo a las leyes y costumbres de la guerra terrestre: "Mientras que se forma un Código más completo de las leyes de la guerra, las Altas Partes Contratantes juzgan oportuno declarar que, en los casos no comprendidos en las disposiciones reglamentarias adoptadas por ellas, las poblaciones y los beligerantes permanecen bajo la garantía y el régimen de los principios del Derecho de Gentes preconizados por los usos establecidos entre las naciones civilizadas, por las leyes de la humanidad y por las exigencias de la conciencia pública".

[14] En esta línea, el *International Committee for Robot Arms Control* promovió en 2012 una petición pública (https://www.icrac.net/the-scientists-call), en 2013 un grupo de ONGs lanzó la campaña *Stop Killer Robots* (https://www.stopkillerrobots.org) y en 2015 miles de investigadores participantes en la *International Joint Conference on Artificial Intelligence*, publicaron una carta abierta solicitando la prohibición del desarrollo, la producción y la venta de armas autónomas.

[15] Concordaría con esta postura Harari cuando dice que los "ordenadores programados con algoritmos éticos podrían someterse con mucha mayor facilidad a los últimos fallos del Tribunal penal internacional" (Harari, 2016, 340).

prohibición absoluta sería preferible una moratoria, al menos hasta que pueda darse un acuerdo más amplio sobre la definición de lo que se pretende disciplinar porque, además, una prohibición preventiva desatendería el imperativo moral a usar la tecnología siempre que sirva para reducir los errores y las atrocidades humanas. Los indicios de la presunta superioridad de los LAWS como actores y observadores 'éticos' provendrían: de la ausencia de emociones que puedan ofuscar el juicio en situaciones que los soldados humanos viven bajo presión; de la no necesidad que experimentan estos sistemas de autoprotegerse; de la capacidad de integrar de forma más completa y rápida la información necesaria antes de responder con un grado letal de fuerza. Claro que, según esta postura, los sistemas autónomos no deberían ser utilizados como remplazo de los soldados, sino para acompañarlos en el campo de batalla (Arkin, 2015, 46-47).

Conviene no olvidar, en todo caso, las debilidades a las que está expuesto el derecho internacional en cuanto a su eficacia. Las incertidumbres arrojadas por la tecnología ya operativa y la inteligencia artificial militar que está llegando exponen la ya frágil operatividad de las normas jurídicas del *ius in bello* a un colapso de grandes proporciones. Las opacas entrañas de los algoritmos 'inteligentes', sumadas a la dificultad para determinar hoy el paradigma jurídico de referencia en el ámbito bélico, aumentan exponencialmente la capacidad de los sistemas autónomos "de producir latencia, de paralizar la eficacia de las normas jurídicas" (Ruschi, 2017, 55) en un entramado como el derecho internacional cuyas ya de por sí flojas mallas legales pueden deshilacharse hasta perder su vigor residual. Siguen pues sobre la mesa algunas cuestiones de hondo calado: ¿las normas internacionales existentes son suficientes para encarar los desafíos puestos por las nuevas tecnologías convergentes? ¿Sería mejor restringir su empleo y desarrollo científico a determinadas situaciones en lugar de una prohibición absoluta, o adoptar un principio de precaución más radical? ¿En definitiva, convendrá esperar y ver lo que se avecina o vislumbrar el futuro desde tan cerca será demasiado tarde? El derecho está más acostumbrado a tratar problemas sociales visibles que a prever los conflictos del futuro y posiblemente una de las notas características del derecho moderno sea la de intervenir cuando ciertos tipos de controversia no han podido recibir un tratamiento institucionalizado. Sin embargo, aunque se han destacado unos cuantos puntos de ruptura entre el pasado, el presente y el futuro con respecto al impacto de la inteligencia artificial, el problema de los efectos que la tecnología provoca en el desarrollo de la guerra no cambia en un aspecto esencial: la necesidad de reducir las consecuencias más destructivas de los conflictos armados.

En realidad, existen normas que buscan limitar el uso de nuevas armas, medios y métodos de guerra: el art. 36 del Protocolo I adicional a los Convenios

de Ginebra obliga a los Estados "a determinar si su empleo, en ciertas condiciones o en todas las circunstancias, estaría prohibido por el presente Protocolo o por cualquier otra norma aplicable (…)". Que la previsión haya sido más o menos desatendida no modifica la novedad del problema arrojado por los LAWS.

Hay al menos dos circunstancias que ponen en jaque a los sistemas jurídicos e impiden aplicar el concepto de responsabilidad asentado en la ciencia jurídica tradicional: una es la paulatina desaparición de la intervención humana en la cadena de actuaciones que determinan el nexo de imputación entre la fuente de la acción y su resultado. La otra tiene que ver con la imprevisibilidad de las decisiones tomadas por los sistemas autónomos guiados por algoritmos cuyos 'enlaces de razonamiento' acaban siendo inaccesibles incluso para quienes los han programado. De hecho, una de las características de los sistemas de inteligencia artificial es que pueden ser programados para comportarse de manera irracional, creando "una metarregla que le indique un cambio de rumbo", siendo este "un rasgo en el que sobresale especialmente el aprendizaje automático" (du Sautoy, 2020, 21). En este aprendizaje autónomo, los algoritmos aprenden de sus fallos atesorando 'lecciones' cuyas explicaciones no han sido introducidas 'desde arriba' por sus programadores, sino que son, por así decirlo, harina de su propio costal[16]. Se perfeccionan a partir de sus errores y lo logran porque se puede crear un algoritmo capaz de plantearse nuevas preguntas si ve que algo no funciona como debería. Todo esto es posible hoy gracias a la inmensa cantidad de datos disponibles que les permiten ejercitarse a una velocidad impensable para el ser humano, ya que en realidad la tecnología que utilizan los sistemas de inteligencia artificial no es tan nueva. Se remonta a sus inicios, a los años cincuenta, cuando se ideó el *perceptrón*, una versión artificial del proceso inferencial de las redes neuronales, en el que una puerta lógica asigna un peso relativo a los datos que influyen en las respuestas a ciertas preguntas relevantes para tomar una decisión. Tal como hacemos los humanos a través de nuestro cerebro, nuestras neuronas se activan en diferentes direcciones según los datos -las señales- que reciben del entorno, así el perceptrón aprendería a imitar el comportamiento cerebral asignando determinados valores a las variables que pueden influir en la decisión y afinando esa ponderación después de cada respuesta equivocada. Esas

[16] En la primera parte de su magnífico libro sobre la creatividad de la inteligencia artificial, cuyo descubrimiento agradezco a la Profa. Mercedes Fuertes, el matemático Marcus du Sautoy ilustra como a partir de DeepMind y AlphaGo ocurrió algo decisivo, que "un algoritmo, construido basándose en un programa que aprende de sus propios fallos, hizo algo nuevo que descolocó a sus creadores y que tuvo un valor increíble. Este algoritmo ganó un juego que, según muchos creían, una máquina no estaba preparada para dominar, ya que era un juego para el que se requiere creatividad" (du Sautoy, 2020, 26).

neuronas artificiales no se han podido plantear hasta una época más reciente porque el combustible que esas redes neuronales precisan para producir resultados son los datos. Y la posibilidad de tratar una cantidad inimaginable de información por parte de los sistemas de inteligencia artificial es lo que ha cambiado exponencialmente en los últimos años y que permite entrenarlos para que lleguen a acertar en la imitación del comportamiento humano (du Sautoy, 2020, 87 ss.). ¿Tal vez necesitaríamos un modelo de *perceptrón normativo* para aprender a analizar los innumerables dilemas surgidos a raíz de la evolución de la inteligencia artificial, usando las normas jurídicas como datos principales para las neuronas artificiales que toman decisiones en el contexto de la guerra y el derecho humanitario? ¿Con un entrenamiento basado en fallos, pero, a la vez, con un optimista horizonte de esperanza en una continua mejora? Alguien tiene que introducir los datos; por tanto, no habría desaparición del papel humano.

Sin embargo, la cuestión principal es la de cómo lograr que el empleo de los sistemas de armamento autónomo pueda garantizar el respeto, además del principio de responsabilidad al que aludía antes, de los principios de distinción, proporcionalidad, necesidad militar, precaución general, etc. Episodios como el de Zemari Ahmadi atestiguan el carácter problemático de exigir a la tecnología bélica la cautela que el derecho internacional impone ante las dudas y la capacidad de adaptarse a las circunstancias del contexto en tiempo real. El robot autónomo debería estar en condiciones de distinguir a un combatiente enemigo que se acerca con intención clara de arrojar un golpe de un enemigo que se acerca a pecho descubierto para rendirse (Balistreri, 2017, 411). La dificultad de interpretar indicios no verbales, contextuales, de la intención humana afecta no solo al principio de distinción, fundamental en el *ius belli*, sino a otro igualmente importante, como es el de proporcionalidad: si un LAWS ha de cumplir una orden de identificar y disparar contra un tanque enemigo, es decir, un objetivo militarmente legítimo, ¿cómo podrá ese robot alterar el significado de ese mandato cuando el tanque se encuentre pegado a una escuela repleta de niños?[17]. Saber analizar una marea infinita de datos y poder hacerlo a una velocidad supersónica, no significa que los sistemas de inteligencia artificial sepan aplicar el sentido común ni que sepan rendir cuenta discursivamente de sus acciones. Para ello necesitarían disponer de una conciencia. Y las interfaces cerebro-máquina no han llegado a tanto. Por decirlo con una imagen literaria empleada por Ian McEwan en *Máquinas como yo y gente como vosotros*, ¿quién va a escribir el algoritmo de la mentira piadosa encaminada a evitar el sonrojo de un amigo?

[17] Sobre la crisis, conceptual y normativa, agudizada por la entrada en los escenarios bélicos de la inteligencia artificial, véase, entre lo más reciente, Aldave, 2022, 193 ss.

Puede sonar siniestro hablar de principios como el de proporcionalidad o de distinción cuando la guerra, a partir de la época nuclear, se ha vuelto inconmensurable desde un punto de vista ético y jurídico, pero ahora mismo no se me ocurre nada más funcional para la reducción de esa desmesura existente entre las posibilidades acarreadas por las tecnologías basadas en la inteligencia artificial y los límites que debemos impedir cruzar para evitar las consecuencias más atroces de la guerra o incluso la completa *algoritmización* de los conflictos armados. Las armas robóticas se suelen dividir en tres categorías según el grado de implicación humana en sus actividades: las *Human* in-the *Loop Weapons*, que pueden seleccionar objetivos y actuar sobre ellos solo a través de un comando humano; las *Human* on-the *Loop Weapons*, que pueden seleccionar y actuar bajo una supervisión humana que puede invalidar la acción del robot; y las *Human* out-of-the *Loop Weapons*, armas que son capaces de seleccionar blancos y atacarlos sin ningún tipo de intervención o input humano (Human Rights Watch, 2012).

La inteligencia artificial es obviamente muy necesaria porque permite hacer muchas cosas y hacerlas mejor, a pesar de que inevitablemente acabe por reproducir ciertos sesgos humanos, pero ello no significa que los programas basados en ella tomen mejores decisiones desde un punto de vista *latu sensu* ético, como por ejemplo si el coche autónomo debería atropellar a la abuela para salvar al nieto o viceversa. Se trata de una lección muy antigua: el mito de Prometeo enseña que el mayor conocimiento 'científico' del mundo (de la *physis*) no conduce a mejores resultados sociales si no está acompañado del conocimiento 'político' (del *nomos*), porque sin el sentido del pudor y la justicia el dominio de la técnica produce efectos incontrolables. Y esa facultad de discernir una idea de lo correcto desaparece si no se sustenta en cadenas racionales que solo afloran *fronéticamente*. Es decir, a través de un intercambio contextual de argumentos en el cual los 'entes' involucrados saben dar cuenta del juicio de ponderación utilizado. En un escenario *out-of-the Loop* esa *accountability* que solo entenderíamos vinculada a valores y emociones humanas se antoja, de momento, impraticable.

Rebus sic stantibus, por tanto, probablemente sea preferible que el poder de decidir asuntos tan relevantes como *quién es el enemigo* dependa de un juicio humano (carbónico) más que de una 'caja negra' algorítmica (silícica). Más que nada porque, a falta de verdades absolutas, el juicio humano parece de momento el único posible de ser sometido al principio del contradictorio, al *audi alteram partem* y, por tanto, el único susceptible de ser argumentado y contra-argumentado. Por muy inteligente que sea esa 'caja negra'.

7. BIBLIOGRAFÍA

Aldave, Ana (2017), *La Guerra Global Contra el Terrorismo. Un análisis de la crisis del Derecho Internacional antes y después del 11-S*, Tirant lo Blanch, Valencia.

— (2022), "Drones, terrorismo e inteligencia artificial: una aproximación a la crisis del paradigma normativo de la guerra", en: Campione, Roger, Filippo Ruschi, Filippo, Ana Aldave (eds.), *Al borde del abismo. Guerra, derecho y tecnología*, Valencia, Tirant lo Blanch, 165-201.

Arkin, Ronald (2015), "The Case for Banning Killer Robots. Counterpoint", en: *Communications of the ACM* 58, 12, 43-44.

Balistreri, Maurizio (2017), "Robot killer. La rivoluzione robotica nella guerra e le questioni morali", en: *Etica & Politica / Ethics & Politics* XIX, 2, 405-430.

Barcellona, Pietro (1996), *El individualismo propietario*, Trotta, Madrid.

Bobbio, Norberto (1992), *El problema de la guerra y las vías de la paz*, Gedisa, Barcelona.

— (1997), *Autobiografia*, ed. de A. Papuzzi, Laterza, Roma-Bari.

Campione, Roger (2005), *La teoría social de Anthony Giddens. Una lectura crítica desde la teoría jurídica*, Dykinson, Madrid.

— (2020), *La plausibilidad del derecho en la era de la inteligencia artificial. Filosofía carbónica y filosofía silícica del derecho*, Dykinson, Madrid.

Chabod, Federico (1967), *Scritti sul Rinascimento*, Einaudi, Turín.

Corradini, Domenico (1995), "Filosofia del conflitto", en: *Rivista Internazionale di Filosofia del Diritto* 1, 9-42.

Du Sautoy Marcus (2020), *Programados para crear. Cómo está aprendiendo a escribir, pintar y pensar la inteligencia artificial*, Acantilado, Barcelona.

Ferrajoli, Luigi (2004), *Las razones jurídicas del pacifismo*, Trotta, Madrid.

— (2007), *Principia iuris. Teoria del diritto e della democrazia. 2. Teoria della democrazia*, Laterza, Roma-Bari.

Goose, Stephen (2015), "The Case for Banning Killer Robots. Point", en: *Communications of the ACM* 58, 12, 43-45.

Kelsen, Hans (1952), *Principles of International Law*, Rinehart & Co, Nueva York.

Harari, Yuval Noah (2016), *Homo Deus. Breve historia del mañana*, Debate, Barcelona.

Hart, Herbert Lionel Adolphus (2012), *The Concept of Law. With a Poscript edited by P.A. Bulloch and J. Raz. And with an Introduction and Notes by L. Green*, 3ª ed., Oxford, Oxford University Press.

Human Rights Watch (2012), *Losing Humanity. The Case against Killer Robots*, International Human Rights Clinic, disponible en https://www.hrw.org/sites/default/files/reports/arms1112_ForUpload.pdf

Machiavelli, Niccolò (1949), *Il príncipe*, en Id., *Tutte le opere*, ed. de F. Flora y C. Cordié, Mondadori, Milán.

Miglio, Gianfranco (2016), *Guerra, pace, diritto*, con un ensayo de Cacciari, Massimo, "La nuova guerra", Editrice La Scuola, Brescia.

Noël, Jean-Christphe (2020), "La inteligencia artificial y el future de la guerra", en: *La Vanguardia Dossier* 77, 64-75.

Parlamento Europeo (2022), *Inteligencia artificial en la era digital*, Resolución del Parlamento Europeo, de 3 de mayo de 2022, sobre la inteligencia artificial en la era digital (2020/2266(INI)), P9_TA(2022)0140.

Rodríguez Fouz, Marta (2022), "El concepto de enemigo y su transformación en los nuevos escenarios bélicos", en: Campione, Roger/Ruschi, Filippo/Aldave, Ana (eds.), *Al borde del abismo. Guerra, derecho y tecnología*, Valencia, Tirant lo Blanch, 25-61.

Ross, Alf (1947), *A Textbook of International Law: general part*, Longmans, London, Green & Co.

Ruschi, Filippo (2017), "El derecho, la guerra y la 'técnica desatada'. Consideraciones acerca de la drone warfare", en: Campione, Roger/Ruschi, Filippo/Aldave, Ana (eds.), *Al borde del abismo. Guerra, derecho y tecnología*, Valencia, Tirant lo Blanch, 45-76.

Sacco, Rodolfo (2007), *Antropologia giuridica*, Il Mulino, Bolonia.

Sánchez Ferlosio, Rafael (2016), *Ensayos 3. Babel contra Babel. Asuntos internacionales. Sobre la guerra. Apuntes de polemología*, Debate, Madrid.

Scalini, Mario (2001), "Tecniche e tecnologie nelle guerre d'Italia", en: Scalini, Mario (ed.), *Giovanni delle Bande Nere*, Milán, Silvana Editoriale, 103-147.

Tamburrini, Guglielmo (2020), *Etica delle macchine. Dilemmi morali per robotica e intelligenza artificiale*, Carocci, Roma.

Tegman, Max (2018), *Vida 3.0. Qué significa ser humano en la era de la inteligencia artificial*, Taurus, Madrid.

Tilly, Charles (1975), "Reflections on the History of European State Making", en: Id. (ed.), *The Formation of National-State in Western Europe*, Princeton University Press.

United Nations, General Assembly, A/65/321, *Report of the Special Rapporteur of the Human Rights Council on extrajudicial, summary or arbitrary executions*, Philip Alston, 23rd august 2010, https://documentsdds-ny.un.org/ /UNDOC/GEN/N10/492/39/PDF/N1049239.pdf?OpenElement

Zagrebelsky, Gustavo (2022), "Quanto sono pericolosi i valori maneggiati dai potente della Terra", en: *La Repubblica*, 13 de abril.

Zolo, Danilo (1998), *I signori della pace*, Carocci Roma.

CAPÍTULO XII

LA JUSTICIA PREDICTIVA: TRES POSIBLES USOS EN LA PRÁCTICA JURÍDICA

MIGUEL DE ASÍS PULIDO
Doctorando del Programa de Doctorado en Derecho y Ciencias Sociales de la UNED
mdeasis@der.uned.es

1. DATAÍSMO Y ALGOCRACIA

En mayo de 2018, el Ministerio de Justicia de España comenzó a utilizar técnicas de gestión de datos para auditar el estado de nuestra justicia, al tiempo que confirmaba su apuesta por la introducción en los próximos años de tecnologías de Inteligencia Artificial y procesamiento del lenguaje natural en el proceso (Bueno de Mata, 2020, 20). En 2020, la Comisión Europea recogía en el apartado 6 de su *Study on the use of innovative technologies in the justice field* una serie de proyectos de los países miembros conducentes a incluir las nuevas tecnologías en el campo de la justicia (Comisión Europea, 2020a). Respecto a España, el documento enumeraba siete proyectos (tres impulsados por el Ministerio de Justicia y cuatro por el Centro de Documentación Judicial -CENDOJ-) que todavía estaban en fase de desarrollo o simplemente habían sido planteados. Estos proyectos comportan tecnologías basadas en Inteligencia Artificial, y consisten en herramientas de transcripción automática de archivos, clasificación de documentos, búsqueda de archivos, seudonimización de sentencias e identificación biométrica. En mayo de 2021, el mismo Ministerio presentó el Plan Justicia 2030, con los Anteproyectos de Leyes de Eficiencia (Ley de Eficiencia Digital, Ley de Eficiencia Procesal y Ley de Eficiencia Organizativa del Servicio Público de Justicia) como base legislativa. Con este Plan, el Ministerio de Justicia apuesta definitivamente por la inclusión de las nuevas tecnologías en el proceso, a fin de mejorar la eficiencia de nuestro sistema judicial dentro del respeto a los Derechos y Libertades de los ciudadanos.

Puede que todavía hubiera quienes, a pesar de las noticias habidas y los textos escritos que proliferan diariamente en nuestros días, se negaran a aceptar la realidad del gran cambio tecnológico. La Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de Inteligencia Artificial (Ley de Inteligencia Artificial) de 2021 a nivel europeo, y la aprobación del Plan Justicia 2030 a nivel nacional, obliga a abrir los ojos a todos aquellos que siguieran dormidos y cautivos en el gran sueño analógico. Y con los ojos fuera de sus órbitas, descubrirán, sin haber tenido tiempo para desperezarse, que nos encontramos ya en los albores de una Edad Media de la Era Digital. En esta Era, caracterizada por la celeridad y

radicalidad de sus cambios y evoluciones, habríamos traspasado el umbral de la Edad Antigua, la Digital, para alcanzar su medioevo: la Edad del algoritmo.

Con el acceso a un número cada vez mayor de datos, petróleo del Siglo XXI y gasolina del algoritmo, y con la extensión de esa especie de religión a la que se le ha puesto el nombre de *dataismo* (Brooks, 2013), las relaciones del presente quedan mediadas y dirigidas por los algoritmos. Las redes sociales, los servicios financieros, la salud, el sector empresarial tecnológico, etc. son ejemplos de ámbitos de nuestra sociedad que han incorporado procesos algorítmicos en su funcionamiento, y no estamos lejos de que estos procesos se extiendan y generalicen en la actuación administrativa. La Inteligencia Artificial, tecnología que aúna datos y algoritmos (Comisión Europea, 2020b, 21), algoritmos y datos, está presente en cada vez más facetas de nuestras vidas. Todo ello nos hace evidenciarnos inmersos en una sociedad *algocrática* (Solar Cayón, 2020, 128), en la que se corre el peligro de que los algoritmos sustituyan al *demos* en el gobierno de nuestras relaciones.

Pese a la protección que frente al cambio le otorga su estricta formalidad, el Derecho también está viéndose afectado por esta Era del algoritmo. Como "conjunto de acciones sociales creadoras «de» o reguladas «por» normas que deben establecer un orden justo en un determinado contexto histórico" (Pérez Luño, 2009, 174), el Derecho se ve afectado por este cambio de paradigma social en un doble sentido. Primero, como creador de normas, como instrumento regulador de las relaciones sociales, pues el propio objeto de su regulación está viéndose modificado. Nuevos campos sociales aparecen, y los campos ya existentes se ven afectados por lo digital y la globalización que las tecnologías de la información y la comunicación (TICs) impulsan. El resultado es la adaptación del sistema jurídico a la sociedad tecnológica, con la creación de una nueva generación de derechos (la tercera generación de Derechos Humanos); la revisión de las regulaciones de distintos sectores, como la salud, las finanzas, el transporte, consumidores, la justicia, etc.; y el impulso de nuevas regulaciones: protección de datos, Internet, Inteligencia Artificial, Biotecnologías, robótica, etc.

Pero, al tiempo que la materia del Derecho se ve afectada, también lo hace su forma. Así, vemos como la práctica jurídica, aparentemente inalterable según el testimonio de los últimos siglos, está mudando de piel en la Era tecnológica. La labor de los abogados, de los procuradores, de los jueces y fiscales, de los funcionarios de justicia, de los paralegales y servicios jurídicos de las empresas, la labor de todos ellos está comenzando a dar señales de un *totum revolutum*. Esta tendencia y esta revolución de la práctica jurídica se ha visto potenciada por la crisis provocada por el SARS-CoV-2 (Covid). Los últimos meses nos han hecho testigos de una caprichosa realidad: la justicia de

nuestra época es una justicia abocada a ser electrónica, a ser e-justicia, a estar atravesada por las TICs (Bueno de Mata, 2020, 11). Si esto era así con la llegada de lo digital, el desarrollo de la Inteligencia Artificial jurídica en los últimos años consolida esta tendencia y aquel *totum revolutum*.

A la hora de estudiar cómo afectan las nuevas tecnologías, en particular la Inteligencia Artificial, a la práctica del Derecho, es imprescindible tratar la cuestión con un enfoque realista, holístico y abierto a la complejidad (Solar Cayón, 2022, 8-9). Una forma de llevar a cabo este estudio de manera efectiva y cumpliendo las citadas exigencias podría pasar por acotar nuestro análisis al proceso judicial. Así, este Capítulo se enmarca en una investigación acerca de la justicia predictiva, en la que se pretende sentar unas bases éticas que garanticen el derecho al debido proceso en la Era algorítmica, que habrá de ser entonces denominado como derecho al debido proceso tecnológico (Citron, 2008). Por cuestiones de espacio, aquí nos limitaremos a tratar ciertos aspectos técnicos y alguno de los usos de la justicia predictiva de sentencias, reservando los comentarios éticos para otra ocasión.

Comenzaremos, en el punto 2, con una breve cronología de la aplicación de la Inteligencia Artificial en el proceso. En ella, hablaremos de los escollos que han retrasado dicha inclusión, al tiempo que se explicará cómo se han ido resolviendo. Tras ello, pasaremos a describir en el punto 3 el funcionamiento de la máquina predictiva de sentencias, donde estudiaremos algunas nociones del aprendizaje automático o *machine learning*. Seguidamente, el punto 4 se reservará a la exposición de nuestra postura acerca de la institución del proceso. Esta exposición resultará imprescindible para comprender los posibles usos de la justicia predictiva en nuestra justicia, usos que constituirán el objeto del punto 5. Por último, se formularán las conclusiones a las que se han llegado en este Capítulo.

2. LA INTELIGENCIA ARTIFICIAL JURÍDICA Y EL PROCESO

La Inteligencia Artificial, como hemos dicho, está alterando profundamente ciertos ámbitos del proceso judicial y de los derechos reconocidos en el mismo (De Asís Pulido, 2021, 67-89). Pero, antes de la Inteligencia Artificial, la Era Digital ya dejó su huella en el proceso. De hecho, el proyecto de modernizar la Administración de Justicia lleva desarrollándose lentamente desde la creación de los primeros ordenadores (Bueno de Mata, 2020, 11) y de su correlato jurídico, la Jurimetría, gestada en 1949 de la mano del juez del Tribunal Supremo de Minnesota L. Loevinger. Esta ciencia jurimétrica ha sido definida de diversas maneras: la aproximación al Derecho a través de métodos computacionales (Francesconi, 2022, 148), el uso de computadores en el Derecho (Belloso Martín, 2021, 353), el conjunto de algoritmos que aplican métodos de cálculo y herramientas de ingeniería en el campo jurídico a fin de

realizar pronósticos sobre el fallo de los jueces y tribunales (Solar Cayón, 2019, 124), etc. Con independencia de la definición que queramos otorgarle, lo cierto es que la Jurimetría impulsó la incorporación de la informática al Derecho y, con ello, la creación de nuevas herramientas procesales. Por su parte, este impulso significó la investigación sobre las posibles aplicaciones de la Inteligencia Artificial al campo jurídico.

A finales de la década de los ochenta, tras un primer invierno de la Inteligencia Artificial, comenzaron a surgir nuevos sistemas expertos, también conocidos como Knowledge Based Systems -KBS-, que a partir de una base de datos de conocimiento y unas reglas lógicas elaboradas por profesionales del Derecho, otorgaban respuestas jurídicas sobre casos concretos. Sin embargo, la imposibilidad de incluir en un sistema lógico todos los matices de la realidad y el Derecho, unido a la insuficiencia por aquel entonces de la capacidad de procesamiento y de conocimiento necesarias para si quiera intentarlo, provocaron un segundo invierno en la Inteligencia Artificial jurídica.

No fue hasta entrado el siglo XXI que esta tecnología volvió a ilusionar a los juristas. La extensión del acceso a la Red, las ingentes cantidades de datos que Internet recolecta y almacena, la decidida apuesta por la estructuración de los mismos, el desarrollo de las técnicas de *machine learning*, de procesamiento de lenguaje natural y de minería de datos han supuesto un nuevo impulso para las tecnologías algorítmicas. Los sistemas expertos del presente superan con creces a los desarrollados en los años ochenta y noventa, pero junto a ellos surgen ahora nuevos sistemas: los sistemas basados en casos, también llamados sistemas basados en datos (*data-driven systems*). Estos sistemas, como veremos, están entrenados con datos: su algoritmo "aprende" automáticamente de ellos (*machine learning*), lo que es lo mismo que decir que el sistema encuentra patrones entre los datos y ajusta el modelo para adaptarse a los mismos.

Intentemos concretar un poco el significado de los *datos* en esto de la Inteligencia Artificial. Ya hemos dicho que aquellos son un elemento primordial de los sistemas algorítmicos. Según la Real Academia de la Lengua Española, un dato, en su tercera acepción, es toda información dispuesta de manera adecuada para su tratamiento por una computadora. En el mundo de la Inteligencia Artificial jurídica, es decir, en el mundo de las herramientas algorítmicas dirigidas a la práctica del derecho, los datos que más nos interesan son los relativos a las normas, jurisprudencia y doctrina. Las bases de datos jurisprudenciales, por ejemplo, ponen a nuestra disposición una gran cantidad de datos jurídicos a través de Internet. En España, por ejemplo, el CENDOJ, órgano técnico del Consejo General del Poder Judicial, se encarga de la publicación oficial y actualizada de jurisprudencia. Por su parte, el artículo 35 del mencionado Anteproyecto de Ley de Eficiencia Digital prevé el desarrollo

de un Portal de Datos de la Administración de Justicia, que deberá proporcionar información procesada y precisa sobre la actividad, carga de trabajo y otros datos relevantes de todos los órganos, servicios y oficinas judiciales y fiscales de España. Todo ello se traduce en una fuente excelente de datos jurídicos con los que alimentar al sistema algorítmico.

Sin embargo, no toda la información sobre normas, jurisprudencia o doctrina que Internet contiene es adecuada para la elaboración de sistemas algorítmicos. Recordemos que la definición de dato nos habla de una disposición adecuada de la información. Y aquí vienen dos problemas fundamentales que tradicionalmente ha arrastrado la Red: su información está escrita en lenguaje natural y la misma carece en gran medida de una estructuración. Respecto al primer problema, en el ámbito del Derecho la cuestión es incluso más grave, pues nos encontramos ante un lenguaje natural muy específico: el lenguaje jurídico (Medvedeva *et al.*, 2020, 242). Para detectar patrones o correlaciones en la información, el sistema antes que nada tendría que ser capaz de procesarla. Pero el algoritmo no entiende el lenguaje natural, pues, además de no tener acceso semántico al lenguaje, solo es capaz de tratar con uno: el maquínico. La solución a este dilema pasa por las técnicas de *Natural Language Processing* (*NLP*), conocidas en castellano como técnicas de procesamiento de lenguaje natural (Medvedeva *et al.*, 2022, 3). Estas técnicas permiten el procesamiento y análisis de datos expresados en lenguaje natural, además de identificar el contexto en el que se presentan y predecir posibles combinaciones de palabras. Es preciso incidir en que este acceso al lenguaje no supone que la máquina lo entienda: la Inteligencia Artificial manipula los símbolos y obtiene correlaciones sintácticas, pero no aprehende ni crea sentido a través de ellos (Luhmann, 2007, 413).

El segundo problema reside en que la información de la Red en muchas ocasiones no está estructurada. En la resolución de este escollo juegan un papel fundamental los metadatos, que podrían definirse como datos que describen el contenido de otros datos, como el autor, el tamaño, la fecha o las palabras clave para reseñarlos. Se trata de una información imprescindible para lograr sistemas de Inteligencia Artificial sólidos y precisos, pues ayudan a la máquina a etiquetar correctamente la información. No es casualidad que el citado Anteproyecto de Ley de Eficiencia Digital, junto con el principio de orientación al dato, establezca la obligación de que la entrada, incorporación y tratamiento de la información se lleve a cabo en forma de metadatos.

La estructuración de la Red y las técnicas de NLP son características de lo que se ha venido denominando como Web 3.0 o Web semántica (Francesconi, 2022, 152). La existencia de esta Web supone un avance en la labor de organización y categorización de la información reproducida en el lenguaje

jurídico, fomentando con ello la aplicación de la Inteligencia Artificial jurídica en el proceso judicial.

Dentro de la Inteligencia Artificial jurídica, en este Capítulo se van a tratar las herramientas de análisis predictivo. Dicho análisis se define como el "área de la minería de datos que combina técnicas de *Big Data*, métodos de aprendizaje automático y modelos estadísticos al objeto de extraer la información existente en los datos y utilizarla para predecir tendencias y patrones de comportamiento" (Solar Cayón, 2019, 125). En el proceso judicial estas herramientas pueden utilizarse para distintos fines: hipótesis de hechos según las pruebas obtenidas por el juez, como las que ofrece el programa STEVIE (Nissan, 2017, 449); predicciones de documentos importantes para el proceso, como las proporcionadas por los sistemas de codificación predictiva; pronósticos del *periculum* para la toma de medidas cautelares, como los obtenidos por los sistemas de evaluación de riesgo, de los cuales COMPAS sería un ejemplo en el ámbito de la reincidencia; pronóstico de resultados de sentencias (*LexMachina)* o de resoluciones alternativas de conflictos, etc.

Algunas de estas herramientas de análisis predictivo son utilizadas en el campo de la justicia predictiva, el cual engloba las técnicas y cálculos dirigidos a pronosticar el sentido de una decisión judicial. Es preciso anotar que la justicia predictiva, entendida en un sentido amplio, no nació con la Inteligencia Artificial. Leibniz ya habló en el siglo XVII de un sistema lógico-matemático que haría predecible la respuesta jurídica (Leibniz, 1989). A principios del siglo XX, Cardozo, un realista americano, definió el Derecho como el "conjunto de principios y dogmas que en una medida razonable de probabilidad podemos predecir como la base de la resolución de controversias pendientes o futuras" (Kelsen, 1979, 198). La citada Jurimetría, en sus orígenes, se limitaba al uso de los computadores y la estadística para la predicción de las decisiones. Es más, la misma posibilidad de justicia predictiva, la capacidad de una persona de saber cuál va a ser la respuesta jurídica ante un hecho, es una exigencia enraizada al Derecho, pues guarda estrecha relación con el principio de seguridad jurídica. No sorprende por ello que, en el marco de la justicia predictiva, Estados Unidos promulgara en el pasado siglo las *sentencing guidelines,* un conjunto de reglas que uniformizaban las decisiones judiciales (Barona Vilar, 2021, 630 y 637).

Si bien es cierto que, como decimos, la justicia predictiva tiene un recorrido más largo que la Inteligencia Artificial, con esta última parece haber dado su salto definitivo a la arena procesal. En este Capítulo nos centraremos en un caso particular de justicia predictiva: la relativa a las sentencias judiciales.

3. SISTEMAS DE PREDICCIÓN DE SENTENCIAS

La superación de la primera ola de Inteligencia Artificial gracias al desarrollo de las tecnologías de *NLP, Big Data, machine learning, data mining* y de la Web semántica ha supuesto el diseño de sistemas que permiten encontrar patrones y correlaciones en los datos inadvertidos por los seres humanos (Susskind, 2020, 315). Estos avances se han producido a costa de un ideal: la creación de máquinas que pensaran como los seres humanos, la creación de una Inteligencia Artificial fuerte que emulase la naturaleza humana. Hoy en día este ideal parece irrealizable: la Inteligencia Artificial que conocemos está basada en tareas específicas (Solar Cayón, 2022, 20-21). Para llevarlas a cabo, el sistema ejecuta una serie de reglas predefinidas, un modelo prestablecido y diseñado por seres humanos que el propio sistema puede ajustar mediante la identificación de nuevos patrones en los datos que se le suministran (*data-driven systems*).

Como ya hemos dicho, los *data-driven systems* son aquellos sistemas que, alimentados por datos, obtienen inferencias de los patrones existentes en los mismos a fin de tomar decisiones o pronosticar sucesos (Hildebrandt, 2018, 3). No resulta de ninguna manera ocioso aclarar aquí que la palabra "predicción" o "*Predict"*, en su sentido estricto, no termina de describir correctamente lo que acaece en los sistemas de Inteligencia Artificial. La extensión de dicha terminología, visible en el propio nombre de estos sistemas (herramientas de justicia predictiva), se debe a su empleo en el ámbito del Procesamiento del Lenguaje Natural y de las técnicas del *machine learning*. Sin embargo, creemos que el término más exacto para referirse a los resultados que ofrecen estos sistemas sería "pronóstico".

Una vez aclarada esta cuestión, hablemos de cómo se consigue el ajuste del modelo en los *data-driven systems*. En primer lugar, en dicho ajuste representan un papel fundamental las técnicas de *machine learning*, que engloban el conjunto de procesos a través de las cuales la máquina "aprende" de los datos. Estas técnicas pueden ser de tres tipos: de aprendizaje supervisado, de aprendizaje no supervisado y de aprendizaje profundo.

Por cuestiones de espacio aquí solo hablaremos de las técnicas de aprendizaje supervisado o *supervised learning*, al ser las que mayor trazabilidad del algoritmo permiten. Existen dos fases en este proceso de aprendizaje: la fase de entrenamiento y la fase de evaluación. En la primera de ellas, se introducen en el sistema una serie de datos de entrada (*inputs*) acompañados de su correspondiente etiqueta (Xiea/Ho/Murphy/Kaiser/Xu/Chen, 2011, 545). En el caso del diseño de herramientas de predicción de sentencias, los datos de entrada se corresponderían al texto de sentencias pasadas o a datos relativos a las mismas, y la etiqueta sería, por ejemplo, si se ha estimado o desestimado las

pretensiones de la demanda. Es preciso contar entonces con un conjunto de datos (o *data set*) debidamente clasificados y etiquetados a fin de entrenar al algoritmo con ellos. Estos datos que, como decimos, pueden tratarse del texto bruto de las sentencias, de una parte de las mismas o de ciertos elementos o características del caso, conforman el *training set.* En la fase de entrenamiento, el sistema identifica patrones de "conexión" entre los *inputs* y las etiquetas y elabora el modelo estadístico que más se ajuste a dichos patrones.

El método de *Support Vector Machine (SVM)* es una de las formas de ajustar el modelo. En el ámbito de la clasificación de sentencias, la utilización de este método consiste en programar a la máquina para que encuentre la ecuación óptima para clasificar los datos (sentencias estimatorias/sentencias desestimatorias) en relación a sus características (Medvedeva *et al.*, 2020, 243). Para que nos sea más fácil comprender a que nos referimos, consideraremos la ecuación como una línea que separa dos conjuntos de datos distribuidos en una gráfica según una serie de características. En esta gráfica, habrá datos que se encuentren en los extremos, situados ahí porque poseerán unas características muy correlacionadas con un sentido estimatorio o desestimatorio de la sentencia. Otros, sin embargo, estarán casi confundidos en posiciones céntricas; se trata, en este último caso, de los vectores soporte. La ecuación del sistema vendrá representada por la línea que separe las dos clases de datos (sentencias estimatorias/sentencias desestimatorias), tomando como referencia los vectores soporte. El modelo se ajustará optimizando el margen de separación entre la línea y dichos vectores.

En lo relativo a las características que se tienen en cuenta, es muy común que en el diseño de herramientas predictivas de sentencias, al tratar en muchas ocasiones con información textual, se atienda a los n-gramas o secuencias de palabras: un unigrama es una palabra, un bigrama es una secuencia de dos palabras, un trigrama una secuencia de tres, un tetragrama una de cuatro, etc. Así, se trata de diseñar a la máquina para que encuentre si existe correlación entre ciertos n-gramas y el resultado de las sentencias. Quizá el sistema identifique que ciertos trigramas sirven de indicadores para ajustar el modelo: el trigrama "el demandado incumplió" puede estar correlacionado con una sentencia estimatoria, mientras que el trigrama "ausencia de irregularidades" podría correlacionarse con una sentencia desestimatoria. En ese caso, estos trigramas funcionarán como predictores o *predictors,* en la medida en que poseerán una correlación positiva o negativa con el resultado (Ronsin/Lampos, 2018, 28). Pero lo importante aquí es que, si sucede que los trigramas son identificados por el sistema como relevantes para determinar el resultado, estos serán los parámetros a través de los cuales se ajustará la ecuación.

Esta selección de datos y parámetros adecuados se hará más sólida y precisa incorporando en la fase de entrenamiento un ciclo de validación (Vadavalasa, 2020, 450-451). En él, se añadirá al sistema otro conjunto de datos, el *validation set*, y se evaluará el rendimiento del modelo, lo que permitirá eliminar datos irrelevantes, evitar el sobreajuste (Russell/Norvic, 2004, 755) y afinar los parámetros.

El algoritmo ya ha sido entrenado para identificar patrones en los datos de entrada (*inputs*) correlacionados con las etiquetas. Ahora es preciso pasar a la segunda fase: la evaluación. En ella se introducirán nuevos datos de entrada (nuevos textos de sentencias), distintos a los datos utilizados en la fase de entrenamiento, pues de lo contrario se sesgaría el modelo (Russell/Norvic, 2004, 752-753). La clave reside en que en esta fase los datos introducidos no vienen acompañados de etiquetas, sino que tendrá que ser el sistema quien las establezca, utilizando para ello el modelo ajustado en la fase de entrenamiento. Al introducir un nuevo texto de sentencia, el sistema ofrecerá el sentido de la decisión (estimatorio o desestimatorio) que le corresponde según la ecuación que ha aprendido. La actuación de la máquina será evaluada mediante el examen de la "precisión" (*accuracy*), referida al porcentaje de resultados que han sido clasificados de manera correcta; o a través del F1-score, medida que resulta de la media armonizada entre la precisión (*precisión*) y la recuperación (*recall*) del sistema. Estos conceptos se refieren al porcentaje de sentencias en las que el resultado ha sido correctamente asignado, en el primer caso, y al porcentaje de casos con un resultado específico que se clasifican correctamente, en el segundo (Medvedeva *et al*., 2022, 4). Pasada la fase de evaluación, el sistema estaría listo. He aquí el milagro de la inteligencia Artificial: la máquina ha creado una ecuación, un nuevo algoritmo a partir del "espacio de cálculo" que se le otorga para llegar al objetivo (Cardon, 2018, 69). La máquina ya es capaz de tomar decisiones y pronosticar eventos futuros en base a los patrones que ha identificado en los casos pasados.

Sin embargo, no todas las herramientas que se diseñan con el fin de predecir el resultado de las sentencias terminan llevando a cabo dicha labor. El origen de unos de los problemas principales de esta tecnología de predicción de sentencias ha recaído desde su surgimiento en los datos que servían de alimento a la máquina. Añadir datos descontextualizados, como la nacionalidad del demandante, el artículo y la ley que considera incumplidos, el nombre del juez, etc. suelen tener el problema de que no permiten captar una imagen integral de los casos; por ello, como ya se ha dicho, se ha tendido a utilizar datos textuales para la alimentación de la máquina, en particular los textos de las sentencias. Recientemente, se han distinguido tres tipos de tareas realizadas por los sistemas textuales de predicción de sentencias: identificación de resultados, clasificación de sentencias y pronóstico de resultados (Medvedeva

et al., 2022, 3). Estas tareas nos obligan a hablar más que de un ámbito de predicción de sentencias, de uno de clasificación, pues solo una de las tres permite obtener un pronóstico como tal de la decisión judicial, a pesar de que dicho pronóstico haya sido el objetivo primordial de todos los estudios y diseñadores de estas herramientas predictivas.

La identificación de resultados se produce cuando el *input* que se suministra a la máquina es *el texto íntegro de la sentencia*, excluyendo el veredicto, pero incluyendo referencias al mismo a lo largo del texto. Con ello, la máquina se limita a identificar el resultado y no a pronosticarlo, porque la información con la que se la alimenta ya viene sesgada por el resultado. Si añades al sistema una sentencia en cuyos Fundamentos de Derecho ya se establece que se he incumplido un contrato, la máquina solo aprenderá a hacer una correlación obvia.

Por su parte, la tarea de clasificación de sentencias es el resultado de alimentar a la máquina con el texto de la sentencia *eliminando el resultado (veredicto) o cualquier referencia al mismo*, pero manteniendo en él cuestiones que solo se conocen una vez dictada la sentencia (hechos probados, normas citadas por el juez, argumentos utilizados por el mismo, etc.). El hecho de que se mantengan estas referencias impide que el programa "aprenda" a pronosticar el resultado de la sentencia, pues para que exista tal pronóstico es necesario que el sistema identifique patrones entre datos anteriores a la sentencia y el resultado. Si el sistema es entrenado con la argumentación judicial, podrá encontrar patrones entre la misma y el resultado, podrá tener un buen rendimiento en la fase de evaluación, pero solo será útil, como la anterior tarea, para obtener inferencias de sentencias ya realizadas. Esta tarea permitiría clasificar la sentencia introducida en un conjunto (estimatorias/desestimatorias, siguiendo nuestro ejemplo), incluso identificar predictores y servir de *checklist* para nuevos casos, pero no serviría para pronosticar resultados de litigios pendientes (Medvedeva *et al.*, 2022, 10).

La tarea de pronóstico de resultados, que es la que aquí nos interesa, solo sería posible si se alimentase a la máquina con datos de la sentencia que se pueden conocer antes de que la misma sea dictada (las alegaciones de las partes, la jerarquía del Tribunal que está resolviendo el caso, ciertos hechos obvios, etc). Se introduciría en el sistema, entonces, solo aquellos fragmentos de la sentencia que cumplieran esta característica. Esto permitiría a la máquina ajustar al modelo de tal manera que pudiera ofrecer el resultado esperado para un caso futuro. Junto a ello, el sistema también sería capaz de detectar predictores en los casos pasados.

En lo que a la práctica del Derecho se refiere, el abanico de posibilidades que surge con esta última tarea es inmenso. Una vez hemos acotado el término

de justicia predictiva a la tarea de pronóstico de sentencias, no huelga decir que la implementación de los usos de esta tecnología en el proceso, cómo lo haremos y hasta qué punto, dependerá del tipo de proceso que queramos que exista en el futuro. Intentaremos ofrecer unas breves notas sobre cómo entendemos dicho proceso, a fin de terminar el Capítulo con una descripción de los posibles usos de la justicia predictiva en el marco judicial expuesto.

4. BREVES NOTAS SOBRE EL PROCESO

El Derecho se podría definir como aquel sistema de normas y principios que regulan la convivencia humana y que la inscriben en un proyecto de Justicia. Este proyecto es aplicado en última instancia por los Tribunales, a través de la percepción, racionalidad y decisión de los jueces (Pérez Luño, 2009, 159): aquellas normas y principios abren al magistrado un marco de posibilidades de respuesta (Martínez García, 2020, 367) dentro del cual, asistido por sus percepciones, su argumentación racional y por el resto de operadores jurídicos, ha de tomar una decisión relativa a una cuestión jurídica controvertida. Una definición genérica del proceso judicial, entonces, sería la que sigue: el proceso es el escenario polifónico donde el juez, en su papel directivo y humano, requerido, auxiliado y controlado por los distintos operadores jurídicos en el marco de una controversia, ha de llevar a la práctica el proyecto de Justicia que subyace al Derecho.

Sin embargo, ocurre que se presentan a menudo en los Tribunales casos que, por su sencillez, no dejan al juez ningún marco de decisión. Coincidiendo con M. Taruffo en que nunca hay dos casos exactamente iguales (Taruffo, 1998, 311), lo cierto es que en el día a día de la práctica judicial sí que se da la existencia de litigios muy parecidos. Hay parcelas del derecho que regulan realidades cuyos matices son bastante reducidos, de tal forma que la respuesta judicial a las mismas suele coincidir en un cúmulo de casos. En ellos, el juez no contaría con ese marco de decisión, ni tendría que recurrir a principios o a argumentos razonables para dictar su respuesta, sino que se limitaría a aplicar la norma o los precedentes a los hechos. Un algoritmo de justicia predictiva que pronosticara la respuesta para un caso podría tomar el papel del juez en dichos litigios, pues su actuación no se diferenciaría de la llevada a cabo por el magistrado. Estos litigios se corresponderían con los casos sencillos, y aunque podría parecer fácil distinguirlos, lo cierto es que existen controversias en su definición. A la hora de estudiar el posible uso de la justicia predictiva en las sentencias, sobre todo a aquellos usos que se refieran a la aplicación del Derecho por parte de estos sistemas, tendremos que aclarar dos cuestiones en relación a estos casos sencillos: primero habrá que encontrar una definición compartida para los mismos, a fin de debatir como sociedad, después, cuáles de estos casos sencillos confiaremos a la máquina para su resolución.

El filósofo del Derecho escocés N. MacCormick caracteriza estos casos sencillos como aquellos en los que no existen problemas de relevancia e interpretación de las normas, la prueba no reviste complejidad y la calificación jurídica está clara (MacCormick, 1978, 227 y ss.). Pero esta no es la única forma de distinguir los casos fáciles de los difíciles. Otra forma sería la que subyace en la diferencia entre los procesos de "justicia equitativa" y los procesos de "justicia codificada" (Re/Solow-Niederman, 2019, 252-255). Los casos de justicia equitativa serían aquellos en los que se aplicarían, además de las normas codificadas, una serie de principios jurídicos, y donde las circunstancias particulares del caso primarían por encima de las reglas generales y abstractas, aunque eso no significara desobedecerlas. Esta justicia equitativa sería aquella en la que el juez tendría que hacer uso de todo su mecanismo hermenéutico, de su razonabilidad, decisión y humanidad. Al contrario, en la "justicia codificada", coincidente con los casos fáciles, se incluirían aquellos litigios para los que se han previsto de antemano todos los escenarios posibles, como ciertos procesos monitorios, alcoholemias en el proceso penal y demandas relativas a clausulas generales de la contratación. Estos casos podrían resultar idóneos para la máquina predictiva.

El Proyecto de Ley de Eficiencia Procesal del Servicio Público de Justicia de 2021 añade a la Ley de Enjuiciamiento Civil el artículo 438.ter, sentando así el concepto de "procedimientos testigos" en el ámbito de las condiciones generales de contratación. El proyecto los define como aquellos que, compartiendo una identidad sustancial de objeto con otros procedimientos en curso, sirven de modelo para la resolución de estos últimos, que quedan suspendidos a la espera de lo que se resuelva en el procedimiento testigo. Este concepto podría aplicarse a lo que llevamos diciendo acerca de los casos sencillos, pues lo que el proyecto de ley está admitiendo es que existen litigios tan estandarizados como para que la resolución señalada en uno de ellos se aplique a todos los demás. Así, tendríamos que la "identidad sustancial del objeto" podría ser otra forma de definir los casos sencillos. Por su parte, las leyes de enjuiciamiento civil y criminal y las reguladoras de las jurisdicciones social y contencioso-administrativa establecen la presencia facultativa de abogado o graduado social en algunos litigios. De nuevo, quizá nos encontramos ante otro criterio que nos sirviera para la distinción entre casos sencillos y difíciles, a pesar de que, en nuestra opinión, el mismo tendría que ser sometido a ciertas revisiones, sobre todo en la jurisdicción social, en la cual es facultativa la presencia del abogado en toda primera instancia.

Estas son las definiciones de proceso y de casos fáciles que proponemos. Como decimos, el uso que se le quiera dar a la Inteligencia Artificial en el proceso dependerá de cómo se posicione la sociedad respecto a estas cuestiones, así como de la forma en la que entendamos la Justicia. Si entendemos al juez

como una máquina que se limita a pronunciar las palabras de la ley, no habrá razones para no sustituir su labor por un sistema automatizado. Si consideramos el proceso como un simple procedimiento de subsunción, en el que todos los casos podrían considerarse sencillos debido a la omnipotencia de la ley, ocurrirá lo mismo. Esa no es nuestra percepción: el juez no solo ejerce una operación deductiva, sino que ha de lidiar con el instrumento abierto, complejo y dinámico que es la norma, aportando su presencia humana en el mismo y comprometiéndose con el proyecto de Justicia latente en el Derecho.

Es cierto que, si se argumenta sin cuidado, existe un potencial peligro en la defensa de un juez humano superior a la máquina, y este consiste en entender la labor judicial como puro arbitrio. Se corre el riesgo entonces de ignorar los mandatos de la seguridad jurídica, pues "cuanto más se critica la Inteligencia Artificial más se defiende la discrecionalidad judicial" (De Asís Roig, 2022, 33-34). La concepción de juez que subyace a nuestra visión del proceso estaría lejos de defender este arbitrio, pues la decisión que ha de tomar el magistrado se enmarca en las posibilidades que le abre el Ordenamiento Jurídico, aunque dicho Ordenamiento sea entendido como algo más complejo y abierto que el mero texto literal de la ley, como un proyecto de Justicia. Quizá habrá casos en los que decidamos que ese proyecto de Justicia puede ser cumplido por una máquina: debido a sencillez y estandarización, consideraremos que la aplicación de los patrones identificados por el sistema en litigios pasados pueda dar lugar a una resolución justa en el caso actual, sin que sea necesario considerar todos los matices de las circunstancias particulares de la controversia.

5. USOS DE LA JUSTICIA PREDICTIVA EN EL PROCESO

Teniendo en cuenta el concepto de proceso que aquí hemos esbozado, la exposición y el análisis sobre los usos de la justicia predictiva en la práctica judicial nos resultará más fácil de contextualizar. Estos usos pueden dividirse en tres conjuntos: aplicación del Derecho, fiscalización de sentencias y pronóstico de resultados para la estrategia procesal.

5.1. **Aplicación del Derecho**

En primer lugar, la máquina predictiva podría utilizarse para proponer la resolución de un caso, de tal manera que la propuesta del sistema podría aplicarse sin intervención judicial alguna o, al menos, servir como recomendación para la decisión judicial. En el primer caso, hablaríamos de una sustitución del magistrado por la máquina predictiva; en el segundo, el sistema funcionaría como un asistente de la decisión humana (*Decision support system*).

La sustitución del ser humano por la máquina en la decisión sobre el fondo del asunto solo podría darse en aquellos litigios que hemos denominado

como casos fáciles, aquellos litigios simples y reiterativos (Nieva Fenoll, 2018, 117) en los que la complejidad de la información relativa al caso y el grado de predictibilidad permitirían a la máquina hacerse cargo de la resolución (Reiling, 2020, 2). En ellos, el juez usa hoy en día herramientas de cortar y pegar, además de recurrir en mayor medida a sesgos cognitivos o heurísticos. Estos sesgos cognitivos son directrices generales que el ser humano utiliza de manera inconsciente para simplificar la realidad a la hora de decidir, lo que nos aporta una visión sesgada de la misma (Nieva Fenoll, 2018, 24, 44-45). Por si esto fuera poco, el magistrado también utiliza modelos de resolución, que estandarizan aún más las respuestas. Una herramienta de justicia predictiva podría resolver estos casos según las correlaciones encontradas por su modelo, lo que haría más eficiente la justicia, aunque fuese en detrimento de la atención a las circunstancias concretas del caso. Además, la máquina podría estar sujeta a evaluaciones y modificaciones periódicas de los algoritmos con el fin de incluir los cambios legislativos, doctrinales o sociales que le permitieran mantenerse actualizada (Volokh, 2019, 1187-1189). Habría que garantizar entonces la transparencia, explicabilidad y confianza del algoritmo.

En nuestro país, la inclusión de la Inteligencia Artificial en la decisión judicial precisaría de la intervención del poder constituyente. El artículo 117 de la Constitución Española establece que la potestad jurisdiccional corresponde exclusivamente a los Jueces y Magistrados, por lo que el uso de sistemas automatizados de sentencias para los casos sencillos tendría que ser precedido de una reforma constitucional, precedida a su vez del debate social sobre los casos sencillos mencionado *supra.* Así mismo, el artículo 2 de la Ley Orgánica del Poder Judicial, que se pretende modificar en el mencionado Proyecto de Ley de Eficiencia Organizativa, habla de Juzgados (Jueces en el Proyecto) y Tribunales, por lo que nada dice sobre la posibilidad de un "juez-robot" o "juez-*software*". Sí que se pronuncia sobre esta cuestión el Anteproyecto de Ley de Eficiencia Digital, pues en su artículo 56.2 establece que la actuación automatizada se podrá aplicar en decisiones que no requieran labores de interpretación jurídica, prohibiendo específicamente en su artículo 57 que el borrador documental aportado por la máquina sea vinculante para el juez en las que sí que requieran dicha interpretación. Por tanto, podríamos concluir que el "juez-robot" no está previsto en el diseño de proceso digital por el que apuesta el Plan Justicia 2030, al menos sobre el papel y en lo relativo a las decisiones sobre el fondo del asunto.

Sin embargo, esto no significa que se haya cerrado definitivamente la puerta a estos sistemas. En caso de que en unos años se decidiera implementar estos "jueces-robot", habría de aplicarse entonces lo establecido en el artículo 42 de la Ley 18/2011, de 5 de julio, reguladora del uso de las tecnologías de la información y la comunicación en la Administración de Justicia: "en caso de

actuación automatizada, deberá establecerse previamente por el Comité técnico estatal de la Administración judicial electrónica la definición de las especificaciones, programación, mantenimiento, supervisión y control de calidad y, en su caso, la auditoría del sistema de información y de su código fuente". Es posible que la ejecución de estos sistemas se haga depender en el futuro de que se incorporen a los mismos tecnologías de argumentación jurídica basadas en reglas (*code-driven systems*) para añadir al resultado una motivación, a fin de respetar el derecho a una motivación suficiente reconocido en el artículo 24 CE. Por último, el ser humano, representado en la figura del juez, a pesar de no estar ya presente en el momento de tomar la decisión, todavía habrá de ejercer un doble control a la máquina: el control en el diseño, garantizándose que el juez participe junto con los técnicos en el mismo; el control a posteriori, asegurándose que la respuesta automatizada pueda ser recurrida ante un juez humano (segunda oportunidad).

Estos requisitos nos parecen los adecuados para garantizar el debido proceso en la resolución automatizada. La mayoría de países que han implementado estas tecnologías, sin embargo, han obviado el proceso legislativo previo, primando la eficiencia de las herramientas predictivas sobre los límites ético-jurídicos y el examen social que ha de preceder a su uso. Las Smart Courts chinas, implementadas en los tribunales superiores de Beijing y Hainan, por ejemplo, resuelven litigios mediante el uso de sistemas predictivos relativos al comercio electrónico e internet, sin que haya habido ningún proceso de distinción entre casos sencillos y complejos, ni ningún debate social sobre la cuestión. Otro ejemplo, esta vez en el ámbito penal y contando con más garantías, es la propuesta del Ministerio de Justicia británico, actualmente suspendida, de aplicar estos sistemas en la resolución automatizada a delitos menores muy estandarizados. Su uso se haría depender de que el acusado se declarase culpable, no alegase circunstancias atenuantes y aceptase someterse a este procedimiento. Igualmente suspendida se encontraría la propuesta del gobierno francés en su programa general de reforma de la Administración de Justicia (2018-2020) de implementar una herramienta de Inteligencia Artificial en la resolución de litigios civiles que no sobrepasen los 6.000 euros (Solar Cayón, 2022, 34-35). Otro país donde existe una propuesta de creación de un sistema predictivo de resolución de asuntos sencillos es Estonia. Desde el 2019, esta propuesta incluiría la resolución automatizada de aquellos litigios sencillos cuya cuantía no superase los 7.000 euros. En este caso, la resolución automatizada estaría especialmente sujeta a apelación (Barona Vilar, 2021, 645).

En aquellos litigios que no fuesen considerados sencillos o para los cuales la sociedad no aceptara la resolución automatizada, los pronósticos y la categorización de sentencias de la máquina predictiva podrían servir como auxilio al juez, proponiéndole una solución (*decisión support systems*). Un

ejemplo de estos sistemas serían los "Smart Judges" chinos, aplicados, entre otros lugares, en el municipio con status provincial de Beijing. Los "Smart Judges", tras considerar cuestiones legales y materiales de un caso, proponen una resolución a los Tribunales (Chen/Li, 2020, 17-18). Por su parte, el sistema PROMETEA, desarrollado por la Fiscalía de la Ciudad Autónoma de Buenos Aires, ofrece un modelo de dictamen jurídico al Fiscal para aquellos casos sencillos en los que se solicita amparo del Tribunal Superior de Justicia de esta ciudad, como los casos de amparo habitacional. Este dictamen, aunque no sea vinculante, habrá de ser tenido obligatoriamente en cuenta por dicho Tribunal. Por tanto, aunque no sea un programa utilizado estrictamente por el juez, sí que entraría dentro del concepto de *decisión support system* en la medida en que, en estos procesos, la oficina del fiscal y la oficina judicial colaboran para resolver el caso. Además, PROMETEA es completamente trazable, lo que garantiza la transparencia del algoritmo. Pese a no ser vinculante, en el primer año de implementación el Tribunal siguió la propuesta ofrecida por la máquina en un 92,2% de los casos. Los resultados relativos a la eficiencia de la justicia son muy positivos tras la implementación de este sistema: la velocidad en la elaboración de informes ha crecido en más de un 300% (Estévez/Fillotrani/Linares Lejarraga, 2020, 15 y 33).

Estos sistemas de auxilio judicial podrían combinarse con otras tecnologías de Inteligencia Artificial. Así, a lo largo del proceso podrían utilizarse herramientas que elaboran hipótesis de los hechos acaecidos en el caso según el material probatorio, como STEVIE, programas de predicción de los riesgos de reincidencia, como COMPAS, y terminar con el auxilio de la máquina predictiva en la elaboración de la sentencia. Esto daría lugar a un sistema integral de *Decision Support*, que sirviera de auxilio al juez durante todo el proceso y que, como ocurría con los "jueces-robot", permitiera incorporar tecnologías de argumentación jurídica para añadir una motivación. El programa Xiao Zhi, implementado en la Corte Suprema Popular de China para litigios relativos a préstamos financieros, sería uno de estos sistemas integrales: esta máquina organiza los eventos del proceso, analiza la presentación de los casos en lo relativo a su admisibilidad, resume los puntos en los que las partes están en desacuerdo, ayuda en la evaluación de las pruebas y crea propuestas de resoluciones judiciales (Chen/Li, 2020, 15). En estos sistemas integrales, si se quisiera respetar el debido proceso, la decisión sobre la resolución le correspondería al juez, pudiendo este último apoyarse en los resultados de la máquina siempre que lo argumentase en su motivación.

El problema que surge con esos *decision support systems* es que "la frontera entre algoritmos de auxilio al juez y algoritmos que sugieren una decisión puede ser muy sutil" (Belloso Martín, 2021, 354). Así, en aquellos casos para los que no autoricemos la respuesta automatizada será necesario asegurar

el derecho del juez de separarse de la decisión propuesta por el sistema. Este derecho tendrá que ir acompañado de una serie de medidas si se quiere superar el sesgo de automatización, consistente en la cesión de la decisión al sistema automatizado, al considerarlo más neutral y menos sujeto al fallo (argumento de autoridad); por miedo a las consecuencias que tendría en la opinión pública un error en caso de haberse separado de la propuesta maquínica (rechazo a las represalias); o por reducir el esfuerzo que supone la toma de decisión (dejadez cognitiva). Este sesgo, lejos de eliminar todo error por descartar de la ecuación la falible decisión humana, provocaría nuevos errores, como el producido por la omisión de la herramienta de propuestas de decisión acertadas o el que surge de aquellos consejos de la máquina que terminan evidenciándose como propuestas de decisión no acertadas (Skitka/Mosier/Burdick, 1999, 993). Por todo ello, la aplicación de estos sistemas en el proceso debería ir acompañada de una formación a la par tecnológica e iusfilosófica de los jueces y de la sociedad entera, que dé cuenta de los límites de la Inteligencia Artificial y de los fundamentos axiológicos de nuestro Derecho y del proceso judicial. Con ello, se pretendería que los jueces se centraran de lleno en su labor primordial, que, como hemos dicho, consiste en llevar a la práctica en las circunstancias concretas el proyecto de Justicia que subyace al Derecho.

Pero la decisión judicial sobre el fondo del asunto no es la única decisión que podría automatizarse mediante las tecnologías predictivas de sentencias. De hecho, en ese sistema integral de *decisión support* que hemos planteado, la herramienta predictiva alimentada con sentencias, además de utilizarse para auxiliar al juez a la hora de resolver el caso, podría haberse aplicado con anterioridad en otros dos momentos: en la decisión sobre la admisibilidad y la preferencia de la demanda o denuncia y en la decisión sobre las normas, jurisprudencia y documentos aplicables al caso (codificación predictiva).

En lo relativo a la admisibilidad, los criterios utilizados por el letrado de la Administración de Justicia y por el juez para tomar la decisión se basan en una serie de requisitos de admisión tasados legalmente para la demanda o la querella, que sirven como primer filtro *in limine litis.* Un sistema predictivo de sentencias podría calcular la probabilidad de éxito de una demanda o querella y, en base a ello, admitirla, inadmitirla o rechazarla. Recordemos que lo que ha servido de alimento para el entrenamiento de estos sistemas no han sido decisiones sobre la admisión o inadmisión, sino textos de sentencias. Pero, realmente, esto no supondría un problema: lo que hace la máquina predictiva es identificar patrones repetidos en decisiones pasadas, y, en su tarea de pronóstico, permite hacerse una idea de las probabilidades que tiene un caso de obtener una sentencia estimatoria en base a los datos que se conocen antes de dictar sentencia. Además, podrían añadirse a estos sistemas un algoritmo

code-driven que tuviera en cuenta, además de los precedentes, los requisitos legales tasados.

Como ya hemos dicho, un ejemplo de herramienta que ayuda en la decisión sobre la admisión de la demanda es el sistema integral de respuesta judicial Xiao Zhi, utilizado en la corte Suprema Popular China (Chen/Li, 2020, 26). En particular, este sistema tiene un problema: al basarse en decisiones pasadas, la herramienta perpetúa ciertos criterios de admisión injustos, pues en China tienden a ser rechazadas por los Tribunales aquellas demandas colectivas que se consideren desestabilizadores del poder del partido y la armonía social (Chen/Li, 2020, 41-44). Sin embargo, lo cierto es que este problema no tiene su origen en la máquina, sino que proviene de la práctica misma del sistema jurídico chino, que la Inteligencia Artificial reproduce.

Por su parte, el orden de preferencia de los asuntos consiste en el orden en el que se señalan aquellos asuntos que hayan sido admitidos. La fijación de los criterios para este orden suele ser competencia de los jueces o los Presidentes de Sala o Sección, pero hay excepciones, como el criterio de urgencia introducido por el Real Decreto Ley 16/2020 tras la crisis del Covid y que se aplicaría hasta el 31 de diciembre de 2020. Un sistema de justicia predictiva que identificara predictores en base a sentencias pasadas podría detectar en nuevos casos indicios de que existe un interés para resolverlos con preferencia. Estos indicios podrían consistir, por ejemplo, en frases que se correlacionaran con incumplimientos graves de ciertos derechos o con casos en los que la víctima fuera una persona vulnerable. En Colombia, desde 2020 está en funcionamiento el proyecto PretorIA, un sistema de Inteligencia Artificial predictiva que establece prioridades entre los casos que llegan a la Corte Constitucional de este país. Entrenado con sentencias, el programa es capaz de identificar información relevante en nuevos recursos en cuestión de segundos (Solar Cayón, 2020, 127), incluyendo predictores que indican la preferencia de ciertos casos. La mayoría de las veces estos predictores se correlacionarán con criterios establecidos por el Tribunal, por lo que PretorIA se trata de un sistema de Inteligencia Artificial que combina las tecnologías *data-driven* y *code-driven*.

Por último, la justicia predictiva de sentencias podría utilizarse para la selección del material normativo y jurisprudencial relevante en el litigio. A este uso de la tecnología se le conoce como codificación predictiva (Solar Cayón, 2019, 145). Una vez admitida la demanda o querella y ya habiéndose señalado la vista, un programa de codificación predictiva podría ofrecer al juez información sobre las normas, las sentencias y los documentos más citados en casos parecidos. Esta información se sumaría al conocimiento jurídico del juez y al que le proporcionan otras herramientas tecnológicas como las bases de datos jurídicas. El programa *Similar Case Intelligent Recommendation System*,

desarrollado en China y disponible para los jueces a través de la *China Justice Big Data Service Platform*, recomienda litigios similares al caso que se pretende resolver, basándose en predictores relativos a los hechos, las normas citadas o la naturaleza de la disputa (Chen/Li, 2020, 17).

5.2. Fiscalización de sentencias

Otro uso de la justicia predictiva, distinto al de la aplicación del Derecho, sería el relativo a la fiscalización de sentencias. Con dicha fiscalización nos referimos a la auditoría de las decisiones judiciales en base a los resultados obtenidos por los sistemas predictivos. Esta auditoría puede tener dos objetivos distintos: la fiscalización restrictiva de la discrecionalidad judicial y la fiscalización restrictiva de los sesgos humanos.

Respecto al primer objetivo, los resultados de las predicciones se emplearían para asegurar que los casos similares reciben una misma respuesta. Aquí la decisión la tomaría el juez, por lo que no se utilizaría la máquina en la aplicación del Derecho. Sin embargo, dicha decisión sería comparada a posteriori con la solución propuesta por la herramienta. En ciertos procesos penales chinos, por ejemplo, se ha desarrollado un sistema de "avisos de sentencias fuera de lo normal", que alerta a las autoridades supervisoras cuando una sentencia se ha excedido de unos límites de discrecionalidad establecidos por casos pasados (Chen/Li, 2020, 19). La implementación de estos sistemas en el proceso judicial podría acarrear, como mínimo, dos posibles problemas: su instrumentalización por parte del poder público, lo que iría en contra de la independencia judicial; y la priorización del pasado sobre las circunstancias particulares del caso, lo que ya ocurría con la utilización de esta tecnología en la aplicación del Derecho. Quizá, la solución a dichos problemas pase por limitar la fiscalización restrictiva de la discrecionalidad a aquellos casos sencillos en los que, como sociedad, primemos la justicia codificada. En ese sentido, habrá que elegir qué sistema implementar respecto a aquellos casos: uno que deje la decisión a la máquina mediante la aplicación automatizada del Derecho u otro que controle la discrecionalidad judicial a posteriori a través de la fiscalización automatizada de las sentencias.

Por otro lado, la justicia predictiva también podría emplearse para identificar sesgos en las decisiones judiciales, a fin de luchar contra los mismos y garantizar el derecho a una decisión imparcial. Esto daría lugar a otro tipo de fiscalización: la fiscalización restrictiva de los sesgos humanos. En base a ella, la máquina predictiva se utilizaría para detectar predictores que manifestasen patrones discriminatorios del sistema judicial, influencias emocionales excesivas, influjos ideológicos inapropiados o el uso acrítico de heurísticos del pensamiento.

Ese marco jurídico que el Ordenamiento abre para la toma de una decisión en el proceso no es el único límite que encuentra el magistrado a la hora de enfrentarse al caso. No hay que olvidar que el juez, como cualquier otra persona, se enfrenta al mundo mediado por el lenguaje y circunscrito a unas circunstancias. Dentro de estas últimas se encuentran sus creencias heredadas, sus ideas sobre el funcionamiento del mundo proporcionadas por la experiencia, la pertenencia a una cultura y a una tradición jurídica determinadas, una personalidad concreta, una forma particular de gestionar las emociones, etc. Además, el lenguaje no es lo único que nos media en el conocimiento de la realidad, sino que también lo hace nuestra forma de conocerla. En ella, son muy importantes los heurísticos o sesgos cognitivos de los que hemos hablado en el anterior epígrafe, que, aunque siempre involuntarios, son a veces irracionales o discriminatorios.

Todo ello podría estar sujeto a un análisis de la máquina predictiva. La tarea de categorización de sentencias permitiría detectar distintos predictores para un resultado específico del proceso que indicasen patrones discriminatorios, como el género de las partes, la nacionalidad o el lugar de residencia del demandado o investigado, etc. En Estados Unidos, en lo relativo a la predicción de riesgo de reincidencia, se demostró que el programa COMPAS, al haber sido alimentado con los datos de un sistema judicial históricamente discriminatorio, reproducía una serie de patrones sesgados respecto a la población afroamericana (Solar Cayón, 2020, 158-159). El programa COMPAS no era racista, pero con su funcionamiento estaba sacando a la luz pública los sesgos ocultos del sistema.

A pesar de que no era su objetivo, el caso de COMPAS es una muestra de que la máquina predictiva podría utilizarse para descubrir sesgos y discriminaciones. Sin embargo, es preciso incidir en que esta fiscalización no podría dar lugar a la creación de perfiles de jueces. La relación pública de jueces individuales con una serie de prejuicios o sesgos iría en contra de su derecho a la intimidad, al honor y a la protección de datos. La Comisión Europea para la Eficiencia de la Justicia (CEPEJ), en su Carta ética europea sobre el uso de la inteligencia artificial en los sistemas judiciales y su entorno, muestra reticencia a dichos perfiles (Comisión Europea para la Eficiencia de la Justicia, 2018, 66-67), estableciendo, eso sí, una excepción: se podrán elaborar cuando dicha información sea accesible únicamente para los jueces, con el único fin de evitar la "patronización" de los fallos judiciales (Belloso Martín, 2021, 357). Además, la detección de los sesgos por parte del sistema predictivo tendrá que venir acompañada de un estudio pormenorizado que tuviera en cuenta el contexto. No hay que olvidar que la máquina también tiene sesgos (Belloso Martin, 2021, 371), y que podría ocurrir que aquello que la misma reconoce como sesgo no sea tal en un caso particular (Ronsin/Lampos, 2018, 29). Se nos ocurren tres

candidatos para realizar dicho estudio o investigación: organismos independientes, como el Comité técnico estatal de la Administración judicial electrónica, órgano creado en la Ley 18/2011, de 5 de julio, reguladora del uso de las tecnologías de la información y la comunicación en la Administración de Justicia; por el Ministerio Fiscal, que actualmente ya realiza funciones parecidas; o por la misma magistratura, mediante la formación de un grupo de expertos dentro del Consejo General del Poder Judicial o la asunción de tareas de fiscalización se sesgos por parte de los Tribunales de instancias superiores.

5.3. Pronóstico para la estrategia procesal

Un último uso de la máquina predictiva, fuera ya de la sede judicial, consistiría en el empleo de sus resultados en el diseño de la estrategia procesal. Aunque el abogado seguiría siendo el protagonista en esta tarea, no podemos negar que la Inteligencia Artificial jurídica fomenta la desintermediación de los servicios jurídicos (Solar Cayón, 2019, 201-205). Así, en algunas ocasiones, la máquina predictiva será utilizada directamente por el usuario para conocer las posibilidades de éxito que tiene en el caso, a la vez que recibe de la misma máquina información relevante sobre normas o precedentes que se aplican al mismo, entre otros servicios posibles. Se tratarán, en todo caso, de los litigios sencillos y reiterativos, en los que el poder predictivo de la Inteligencia Artificial poseerá altos niveles de precisión. En los demás casos, y sobre todo en los litigios descritos como difíciles o de "justicia equitativa", la presencia del abogado y del procurador seguirá siendo insoslayable para garantizar el derecho a la defensa de los ciudadanos.

No hay que soslayar el hecho de que, fundamentalmente, la tarea del abogado consiste en la creación de una estrategia de defensa o acusación basada en la predicción de lo que el juez va a resolver en un caso concreto (De Asís Pulido, 2020, 189-190). En ese sentido, la figura tradicional de abogado está siendo protagonista de un cambio sin precedentes. Pensemos en lo revolucionario, respecto a la labor del abogado, de tecnologías como las bases de datos jurídicas, el acceso a Internet o, sobre todo, las nuevas herramientas de Inteligencia Artificial. Un ejemplo de estas últimas, como decimos, son los sistemas de análisis predictivo, que ofrecen información de gran utilidad sobre el caso al jurista. Estos sistemas ofrecen un pronóstico de la respuesta judicial en base a los datos que, tras reunirse con el cliente, el abogado ha podido recabar, lo que le facilitaría la decisión sobre si merece la pena llevar el litigio a los Tribunales y, en su caso, organizar una estrategia procesal que tuviera en cuenta dicho pronóstico. En Estados Unidos, la compañía *Lex Machina* (Solar Cayón, 2019, 133), adquirida por *LexisNexis* en 2015, ofrece servicios de categorización y pronóstico de sentencias en materia de propiedad intelectual y de legislación antitrust, siendo más efectivas en este campo que cualquier

abogado (Susskind, 2017, 186). Otras compañías que ofrecen dichos servicios de análisis predictivo en ciertas materias jurídicas son Wolters Kluwer, a través de la Ley Digital, y Tirant lo Blanch, en su plataforma Tirant Analytics.

Como hemos dicho, este pronóstico de la decisión judicial también podría ser útil para los mismos usuarios, que accederían a predicciones sólidas sobre lo que puede ocurrir en un caso sin necesidad de recurrir a un abogado. Además, los predictores identificados por la máquina podrían servir para realizar una codificación predictiva de las normas, los precedentes o las pruebas más utilizadas en casos similares. Esta información sería de enorme provecho para el abogado, pero también para aquellos usuarios que decidiesen acceder a la justicia sin intervención de letrado o de procurador. Por último, si fuésemos capaces de incorporar a la tecnología predictiva de sentencias una parte de *code-driven* y de sistemas *data-driven* alimentados con demandas o denuncias pasadas, lograríamos desarrollar sistemas integrales de pronóstico, codificación y creación automatizada de documentos jurídicos. Estos sistemas integrales, lejos de sustituir a los abogados, les permitirían centrarse en las cuestiones más importantes y complejas del caso, al tiempo que mejorarían en grado sumo el acceso a la justicia a los ciudadanos.

6. CONCLUSIÓN

Hemos dicho en la Introducción que la materia y la práctica del Derecho están mudando de piel. Su cuerpo, conformado por el respeto al Ordenamiento Jurídico y la búsqueda del orden justo en un determinado contexto histórico, parece aun mantenerse bajo las dos pieles, la muerta y la que está por venir. Muestra de ello son todos los documentos y propuestas de Reglamento que han venido elaborándose en el seno de la Unión Europea en los últimos años, a fin de sentar unos principios éticos de la Inteligencia Artificial (Barona Vilar, 2021; Llano Alonso, 2021; De Asís Roig, 2022). Algunos de estos documentos, como los elaborados por el CEPEJ, se refieren exclusivamente al campo procesal. Pero sería imprudente cantar victoria por ello y avanzar despreocupados hacia un mañana que en el presente se hace incierto.

La Inteligencia Artificial aplicada al proceso tiene como principal objetivo vencer el colapso que sufre nuestra Administración de Justicia, y hacerla más rápida, abierta, eficiente, precisa, transparente y próxima al ciudadano (Bueno de Mata, 2020, 11). En particular, el uso de la justicia predictiva para la aplicación del Derecho reduciría los tiempos de análisis y generación de resoluciones, abarataría los costes del proceso y mejoraría el acceso a la justicia de los ciudadanos, al tiempo que descargaría la oficina judicial de ciertos casos sencillos y permitiría a los jueces centrarse en litigios más complejos, que hemos denominado de "justicia equitativa". La fiscalización de sentencias conllevaría un nuevo control a la actividad judicial, a través del cual se conseguiría detectar

ciertos sesgos o patrones discriminatorios presentes en el sistema jurídico, lo que redundaría en el derecho a la imparcialidad judicial. Por su parte, el acceso a un pronóstico de lo que se va a decidir en un caso concreto supondría una ayuda sin par a los abogados y usuarios en su elaboración de la estrategia procesal.

Sin embargo, ciertos usos de la Inteligencia Artificial, y de la justicia predictiva en particular, podrían terminar imponiendo una nueva concepción de la Justicia, en la que el respeto a los Derechos Humanos y a la dignidad quedaran reemplazados por una lógica de la eficiencia y el beneficio, y donde las circunstancias particulares del caso fueran sustituidas por una aplicación sistemática del pasado en el presente. En esta nueva Justicia, el juez podría verse impelido a seguir la solución propuesta por la máquina, ya fuese por un impulso externo (sesgo de automatización) o por uno externo (un sistema de avisos que fiscalizara sus decisiones y coartara su discrecionalidad). Esto supondría extender el uso de la aplicación automática de los sistemas predictivos más allá de los que se ha denominado *supra* como casos sencillos. Significaría, en definitiva, que el magistrado característico de esta Justicia de la eficiencia volvería a ser mudo; con la diferencia de que esta vez, en lugar de las palabras de la ley, su boca balbucearía el lenguaje codificado del algoritmo.

Otro problema que podría traer esta nueva Justicia sería la vulneración de la intimidad y de la protección de datos de los jueces, con la creación de perfiles de los mismos; perfiles que podrían extenderse también al resto de operadores jurídicos e incluso a las partes. La brecha tecnológica, la sustitución en todo litigio del abogado por la Inteligencia Artificial o la renuncia consolidada a defender aquellos casos que la máquina predictiva pronostique como "imposibles" podría también dañar los principios básicos de igualdad de partes y de derecho a la defensa.

Para prevenir esta posible deriva de la Inteligencia Artificial, el CEPEJ, en su Carta ética de 2018, incluyó la creación de perfiles y la anticipación de las decisiones judiciales dentro de los usos tecnológicos que han de ser estudiados antes de su aplicación definitiva (Comisión Europea para la Eficiencia de la Justicia, 2018, 67). Ya por entonces la misma Carta ética presentaba sus reservas más extremas a los sistemas de fiscalización restrictiva de la discrecionalidad judicial, a los que se refería como sistemas de "normas basadas en la cantidad" (Comisión Europea para la Eficiencia de la Justicia, 2018, 68). Por su parte, la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de Inteligencia Artificial (Ley de Inteligencia Artificial) de 2021, califica en el punto octavo de su Anexo III como de alto riesgo toda Inteligencia Artificial que asista al Juez a la hora de administrar Justicia.

Vemos como, al menos a nivel europeo, se va reaccionando poco a poco a los peligros de la Inteligencia Artificial jurídica, aunque por ahora esta reacción se haya dado casi en exclusiva desde un enfoque ético. En el futuro próximo se hace urgente, en base a lo avanzado desde este enfoque, sentar las bases jurídicas para la inclusión de la Inteligencia Artificial en el proceso judicial. Estas bases pasarán inevitablemente por la "elección" de aquellos casos sencillos en los que queramos aplicar la máquina predictiva como "juez-robot" o "abogado-robot", así como el nivel de "vinculatoriedad" que se le querrá dar a la propuesta maquínica de resolución judicial (Llano Alonso, 2022, 8), elecciones que precisarán de un debate social.

Hemos definido al Derecho, en palabras de A. E. Pérez Luño, como el "conjunto de acciones sociales creadoras «de» o reguladas «por» normas que deben establecer un orden justo en un determinado contexto histórico" (Pérez Luño, 2009, 174). Por otro lado, hemos considerado el proceso judicial, esta vez con palabras propias, como el escenario polifónico donde el juez, en su papel directivo y humano, requerido, auxiliado y controlado en el marco de una controversia, ha de llevar a la práctica el proyecto de Justicia que subyace al Derecho. Pues bien, ese orden justo y ese proyecto de Justicia, en nuestro contexto histórico y social, pasa por la garantía de los Derechos Humanos, radicados en la dignidad de la persona, y habrá que pugnar decididamente por una Era algorítmica que se enmarque dentro de este orden. Por ello, alineándonos con el enfoque basado en el riesgo y con la mirada ética de la tecnología por los que la Unión Europea ha estado trabajando en los últimos años (Comisión Europea, 2020b, 22), creemos que se hace preciso sentar unos principios éticos de la justicia predictiva que se incorporen a la regulación jurídica del proceso, y en particular al derecho fundamental al debido proceso, que habrá de ser tecnológico o no será. Nos limitaremos aquí a nombrar alguno de ellos, reservándonos la profundización sobre los mismos para otro lugar: *respeto a los Derechos Fundamentales y a la autonomía humana; no discriminación; protección y calidad de los datos; precisión, solidez y seguridad del sistema; trasparencia de los algoritmos; imparcialidad; colaboración jurista-técnico en el diseño;* y *control humano del sistema.*

7. BIBLIOGRAFÍA

Barona Vilar, Silvia (2021), *Algoritmización del Derecho y de la Justicia. De la Inteligencia Artificial a la Smart Justice*, Tirant lo Blanch,Valencia.

Belloso Martín, Nuria (2021), "Los desafíos iusfilosóficos de los usos de la Inteligencia Artificial en los sistemas judiciales: a propósito de la decisión judicial robótica *vs.* Decisión judicial humana", en: Belloso Martín, Nuria (ed.), *Sociedad plural y nuevos retos del Derecho*, Pamplona, Thomson Reuters Aranzadi, 327-401.

Brooks, David (2013), "The Philosophy of Data", en: *The New York Times,* 4 de febrero de 2013. https://www.nytimes.com/2013/02/05/opinion/brooks-the-philosophy-of-data.html.

Bueno De Mata, Federico (2020), "Macrodatos, Inteligencia Artificial y proceso: luces y sombras", en: *Revista General de Derecho Procesal* 51, 1-32.

Cardon, Dominique (2018), "The power of algorithms", en: *Pouvoirs* 164, 1, 63-73. https://www.cairn-int.info/article-E_POUV_164_0063--the-power-of-algorithms.htm.

Citron, Danielle Keats (2008), "Technological Due Process", en: *Washington University Law Review* 85, 3, 1249-1313.

Chen, Benjamin Minhao, Li, Zhiyu (2020), "How will technology change the face of Chinese justice?", en: *Columbia Journal of Asian Law* 34, 1, 1-58.

Comisión Europea (2020a), *Study on the use of innovative technologies in the justice field.* https://op.europa.eu/en/publication-detail/-/publication/4fb8e194-f634-11ea-991b-01aa75ed71a1/language-en.

— (2020b), *Libro Blanco sobre la inteligencia artificial: un enfoque europeo orientado a la excelencia y la confianza, Bruselas*, COM (2020) 65 final, 19 de febrero de 2020. https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_es.pdf.

Comisión Europea Para La Eficiencia De La Justicia (2018), *European Ethical Charter on the use of Artificial Intelligence in Judicial Systems and their environment*, Estrasburgo, 3-4 de diciembre de 2018. https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c.

De Asís Roig, Rafael (2022), *Derechos y Tecnologías,* Dykinson, Madrid.

De Asís Pulido, Miguel (2020), "La incidencia de las nuevas tecnologías en el derecho al debido proceso", en: *Ius et Scientia* 6, 2, 186-199. https://revistascientificas.us.es/index.php/ies/article/view/14337/12777.

— (2021), "Derecho al debido proceso e Inteligencia Artificial", en: Llano Alonso, Fernando, Garrido Martín, Joaquín (eds.), *Inteligencia Artificial y Derecho. El jurista ante los retos de la Era Digital*, Pamplona, Thomson Reuters Aranzadi, 67-89.

Estevez, Elsa, Fillotrani, Pablo, Linares Lejarraga, Sebastián (2020), "PROMETEA: Transformando la administración de justicia con herramientas de inteligencia artificial", en: *Banco Interamericano de Desarrollo*. doi:10.18235/0002378.

Francesconi, Enrico (2022), "The winter, the summer and the summer dream of artificial intelligence in law. Presidential address to the 18th International Conference on Artificial Intelligence and Law", en: *Artificial Intelligence and Law* 30, 147-161.

Hildebrandt, Mireille (2018), "Algorithmic Regulation and the Rule of Law", en: *Philosophical Transactions Royal Society*, 1-11. https://royalsocietypublishing.org/doi/full/10.1098/rsta.2017.0355.

Kelsen, Hans (1979), *Teoría general del Derecho y del Estado*, Universidad Nacional Autónoma de México, México.

Leibniz, Gottfried Wilhelm (1989), "*De Arte combinatoria"*, en: Leibniz, Gottfried Wilhelm, *Philosophical Papers and Letters,* 2ª Ed., Dordrecht, Kluwer Academic Publishers.

Llano Alonso, Fernando (2021), "Ethics of Artificial Intelligence in the European Union Legal Framework", en: *Ragion pratica Rivista semestrale* 2, 327-348.

— (2022), "Justicia Digital, algoritmos y Derecho", en: Solar Cayón, José Ignacio, Sánchez Martínez, María Olga (eds.), *El impacto de la Inteligencia Artificial en la teoría y la práctica jurídica*, Madrid, Wolters Kluwer. En prensa. Se ha podido acceder al texto por gentileza del autor.

Luhmann, Niklas (2007), *La sociedad de la sociedad*, trad. de J. Torres Nafarrate, Herder, México.

MacCormick, Neil (1978), *Legal Reasoning and Legal Theory,* Oxford University Press, Oxford.

Martínez García, Jesús Ignacio (2020), "La respuesta jurídica", en: *Anuario de Filosofía del Derecho,* XXXVI, Madrid, BOE, 347-371.

Medvedeva, Masha *et al.* (2020), "Using machine learning to predict decisions of the European Court of Human Rights", en: *Artificial Intelligence and Law* 28, 237-266. https://doi.org/10.1007/s10506-019-09255-y.

— (2022), "Rethinking the field of automatic prediction of court decisions", en: *Artificial Intelligence and Law*, 1-19. https://doi.org/10.1007/s10506-021-09306-3.

Nieva Fenoll, Jordi (2018), *Inteligencia artificial y proceso judicial*, Marcial Pons, Madrid.

Nissan, Ephraim (2017), "Digital technologies and artificial intelligence's present and foreseeable impact on lawyering, judging, policing and law enforcement", en: *AI & Society* 32, 441-464.

Pérez Luño, Antonio Enrique (2009), "¿Qué significa juzgar?", en: *DOXA, Cuadernos de Filosofía del Derecho* 32, 151-176.

Re, Richard M., Solow-Niederman, Alicia (2019), "Developing artificially intelligent justice", en: *Stanford Technology Law Review* 22, 242-289.

Reiling, Dory (2020), "Courts and Artificial Intelligence", en: *International Journal for Court Administration* 11, 2, art. 8, 1-10.

Ronsin, Xavier, Lampos, Vasileios (2018), "In-depth study on the use of AI in judicial systems, notably AI applications processing judicial decision and data", en: EUROPEAN COMMISION FOR THE EFFICIENCY OF JUSTICE, *European Ethical Charter on the Use of artificial Intelligence in Judicial Systems and their environment,* Strasbourg, 13-63.

Russell, Stuart, Norvic, Peter (2004), *Inteligencia Artificial: un enfoque moderno*, Pearson, Madrid.

Skitka, Linda J., Mosier, Kathleen L., Burdick, Mark (1999), "Does Automation Bias Decision-making?", en: *International Journal of Human Computer Studies* 51, 991-1006.

Solar Cayón, José Ignacio (2019), *La Inteligencia Artificial Jurídica, El impacto de la innovación tecnológica en la práctica del Derecho y el mercado de servicios jurídicos*, Aranzadi, Navarra.

— (2020), "Inteligencia Artificial en la Justicia Penal: los sistemas algorítmicos de evaluación de riesgos", en: Solar Cayón, José Ignacio (ed.), *Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho*, Madrid, Editorial Universidad de Alcalá, 125-172.

— (2022), "¿Jueces-robot? Bases para una reflexión realista sobre la aplicación de la inteligencia artificial en la Administración de Justicia", en: Solar Cayón, José Ignacio, Sánchez Martínez, María Olga (eds.), *El impacto de la inteligencia artificial en la teoría y la práctica jurídica*, Madrid, Wolters Kluwer. En prensa. Se ha podido acceder al texto por gentileza del autor.

Susskind, Richard (2020), *Tribunales online y la Justicia del futuro,* trad. esp. GEA Textos, La Ley - Wolters Kluwer, Madrid.

— (2017), *Tomorrow's Lawyers*, Oxford University Press, Oxford.

Taruffo, Michele (1998), "Judicial Decisions and Artificial Intelligence", en: *Artificial Intelligence and Law*, 207-220.

Vadavalasa, Rammohan M. (2020), "Data Validation Process in Machine Learning Pipeline", en: *IJSRD - International Journal for Scientific Research & Development* 8, 4, 449-452.

Volokh, Eugene (2019), "Chief Justice Robots", en: *Duke Law Journal* 68, 1135-1192. https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=3973&context=dlj.

Xiea, Xiaoyuan, Ho, Joshua W. K, Murphy, Christian, Kaiser, Gail, Xu, Baowen, Chen, Tsong Yueh (2011), "Testing and Validating Machine Learning Classifiers by Metamorphic Testing", en: *Journal of Systems and Software* 84, 4, 544-558.

CAPÍTULO XIII

PROTECCIÓN PENAL DE LOS NEURODERECHOS: EL USO DIRECTO DE LAS NEUROTECNOLOGÍAS SOBRE EL SER HUMANO

Mª ISABEL GONZÁLEZ TAPIA
Universidad de Córdoba
fd1gotam@uco.es

1. INTRODUCCIÓN

La relación existente entre la neurociencia y la inteligencia artificial es una relación simbiótica (Savage, 2019). No se entienden en toda su dimensión, la una sin la otra. La inteligencia artificial ha proporcionado a la neurociencia, de una parte, la capacidad computacional necesaria para aprender aspectos muy complejos de la naturaleza del ser humano. Ha aportado la capacidad para gestionar e integrar con sentido una miríada de datos neuronales, genéticos, hormonales, psiquiátricos o psicológicos y conductuales…, interactuando para proporcionar un mejor conocimiento de cómo somos, cómo pensamos o sentimos, de cómo aprendemos, de cómo decidimos y por qué actuamos. También ha permitido, por así decirlo, "leer" lo que está pensando o sintiendo el ser humano en tiempo real y transformarlo en visiones, en lenguaje, en acciones, en emociones…que, incluso, pueden transmitirse a distancia o a dispositivos robóticos… De otra parte, la inteligencia artificial aporta su capacidad predictiva, formulando pronósticos dinámicos sobre el surgimiento y evolución de enfermedades, sobre cómo vamos a reaccionar ante determinados estímulos, sobre qué vamos a decidir, sobre nuestro comportamiento… Del mismo modo, el acceso a *big data* neurobiológicos y el uso combinado de neurotecnologías y de sistemas de inteligencia artificial parece que permitirá en un futuro próximo que también se pueda "escribir" en nuestro cerebro, generando artificialmente estímulos, incluyendo visiones, incluyendo o retirando memorias, generando impulsos motores, generando o reprimiendo emociones…; muchas veces a través de dispositivos que permiten, por ejemplo, percepción sensorial artificial: como la retina de silicio, nariz artificial, cóclea artificial, chip sónar o chips de percepción de movimiento; dispositivos que detectan estados de ánimo y son capaces de incidir en ellos; que detectan y controlan enfermedades en estadios embrionarios… En definitiva: la inteligencia artificial ha permitido a la genética y la neurociencia mejorar de forma radical el conocimiento de cómo somos, de cómo pensamos y decidimos y cómo sentimos; y hacerlo en forma dinámica, permitiendo pronósticos de situaciones y comportamientos futuros y eficacia de los tratamientos mucho más precisos. Y la inteligencia artificial, a partir de ese conocimiento, se ha

desarrollado a imagen y semejanza del cerebro humano, pero con una capacidad cuántica de computación (y multiplicándose), permitiendo la captación de datos neuronales a gran escala (*big data*), la lectura e interpretación de los mismos y la capacidad de hacer predicciones autónomas y de manipularlos o cambiarlos: la capacidad como se ha dicho gráficamente por Yuste y colegas, 2017, de "leer y escribir" nuestro cerebro. Nada menos.

Y eso es algo, simplemente, maravilloso. Basta pensar en las posibilidades que se están abriendo con el uso terapéutico de las neurotecnologías. Por ejemplo, utilizar estimulación cerebral profunda para el tratamiento del Parkinson, la epilepsia, la depresión mayor, trastornos obsesivos compulsivos, adicciones, fobias, trastorno de estrés postraumático o el dolor… Usar herramientas de inteligencia artificial para el diagnóstico precoz de enfermedades o de brotes como la epilepsia (Cook *et al.*, 2013) o la esquizofrenia… O más aún, la utilización de dispositivos de interfaz cerebro-computador que graban la actividad neurológica y, después de un entrenamiento con el usuario, permiten identificar y usar patrones neurológicos asociados con intenciones motoras de éste y controlar a partir de ahí objetos, permitiendo a personas con parálisis o dificultades motoras severas manejar extensiones robóticas (Kögel *et al.*, 2020; Hochberg, 2012, Muelling *et al.*, 2017; entre otros), o hablar únicamente con el pensamiento (p.e. Sellers *et al.*, 2010; Wolpaw *et al.*, 2018) o recuperar memorias en el Alzeimer, o ver a las personas invidentes, o detectar y regular emociones (Stainert *et al.*, 2020)… Hacer cosas con el pensamiento (Steiner *et al.*, 2019). El potencial aplicativo de la inteligencia artificial al servicio del ser humano, por ejemplo, en el campo de la salud, no tiene parangón. Es absolutamente extraordinario. Pero, inevitablemente, ese mismo potencial extraordinario también abre un "portal" hacia nuevas fuentes de riesgo para el ser humano, para nuestra sociedad y para nuestra forma de vida. El reconocimiento de estos riesgos potenciales, derivados posibles "*dobles usos*" de las neurotecnologías (Goering *et al.*, 2021) y de la inteligencia artificial, justifica que el Derecho Penal deba entrar en escena para reflexionar, desde un punto de vista político-criminal, si debe intervenir en este ámbito o no y en qué términos habría de hacerlo.

2. NEUROTECNOLOGÍAS, NUEVOS RIESGOS Y NECESIDAD DE LA INTERVENCIÓN PENAL

Las relaciones o implicaciones que pueden establecerse entre el Derecho Penal y el uso de la inteligencia artificial son muy amplias; dichas relaciones, cuando se plantean desde la interacción entre la inteligencia artificial y la neurociencia, son aún más amplias y con un calado radical. Probablemente, en un futuro más o menos próximo, se va a producir un cambio de paradigma, ante un nuevo contexto socioeconómico digitalizado y virtualizado, en el marco

del "*neurocapitalismo*" de la vigilancia (Cfr. Zuboff, 2020), donde la posibilidad de manipular al ciudadano y a grupos enteros sea cierta y en el que convivan humanos y meta-humanos… Un contexto que parece tener el potencial de alterar sustancialmente los dos pilares sobre los que se ha construido dogmáticamente el Derecho Penal. De una parte, la función preventiva y garantista de nuestro modelo de Derecho Penal, lo que equivale a la función de las penas y de las medidas de seguridad y el juego de los principios limitadores del *ius puniendi*. De otra, las bases mismas de la responsabilidad penal, es decir, de la culpabilidad.

En este capítulo nos vamos a centrar únicamente en los riesgos derivados del uso directo de las neurotecnologías sobre el ser humano, en atención, como se ha dicho, al potencial "doble uso" que de ellas puede hacerse. Nuestro foco se va a situar en el potencial uso pernicioso (efectivo o *razonablemente previsible*) que pudiera hacerse de la neurotecnología, susceptible de lesionar o poner en peligro al ser humano y a nociones básicas de nuestra convivencia, en un modelo de Estado Social y Democrático de Derecho. Y el objetivo va a ser formular un análisis político-criminal siquiera aproximativo de algunos de esos nuevos riesgos, derivados de: 1) la captación y tratamiento de datos neuro-biológicos y del uso la información de salida para ejercer una prevención del delito o un control social no democráticos e ilegítimos; 2) la manipulación o alteración de la autonomía de individuos y grupos; 3) la provocación de daños o peligros a la salud mental de los ciudadanos. Es decir, nos centraremos en los nuevos riesgos y desafíos derivados del uso directo de la neurotecnología sobre el ser humano.

Y la cuestión a debatir no va a ser de prohibiciones sino de *límites y de capacidad* real del Derecho Penal para imponer tales límites, que están siendo reclamados como necesidad urgente, incluso (y, sobre todo) por quienes se hallan en la vanguardia de la investigación acerca del uso de las neutorecnologías directamente sobre el ser humano. Así, por ejemplo, Elon Musk, fundador de *Neuralink*, empresa dedicada al "amejoramiento" cerebral, decía ya en 2014, en la inauguración del MIT Aeronautics and Astronautics department's Centennial Symposium, que necesitamos claros límites normativos, nacionales e internacionales, porque con la inteligencia artificial estamos "convocando al demonio"[1]. Así mismo, el neurobiológico Rafael Yuste, Catedrático de la Universidad de Columbia y líder del proyecto BRAIN, también lleva años advirtiendo de los peligros que puede entrañar el (mal-)uso de las neurotecnologías y de la necesidad de reconocer nuevos *neuroderechos*

[1] Cfr. el discurso de Elon Musk en: https://www.washingtonpost.com/news/innovations/wp/2014/10/24/elon-musk-with-artificial-intelligence-we-are-summoning-the-demon/ (últ. visita 13/12/2021). En cuanto a las declaraciones de Juste y el grupo Morningside: Vid. (Yuste *et al.*, 2017); o https://www.braininitiative.org/, por ejemplo.

que protejan la diversidad, la autonomía y la identidad del ser humano. Así, ya en 2017 reclamaba en la revista *Nature* (Yuste *et al.*, 2017), como representante de un grupo informal interdisciplinar llamado *Morningside* la necesidad de ocuparse ya de los desafíos éticos y jurídicos que plantean las neurotecnologías, dada su potencialidad para exacerbar desigualdades sociales y para brindar oportunidades de manipulación o, incluso, de alterar la propia naturaleza humana. Y recientemente, en una entrevista concedida al periódico *El País*, junto con el ingeniero Darío Gil (Ansede, 2022), afirmaba sin ambages que en un futuro próximo (a partir de 10 años) la sociedad comenzará a dividirse sin remedio entre seres humanos y seres híbridos, mejorados con técnicas invasivas o no, que estarían a disposición del consumidor. Y, precisamente, afirman, la posibilidad de que pueda hacerse un uso comercial de estas neurotecnologías hace aún más perentorio intervenir desde el principio, antes de su implantación social, para asegurar un desarrollo ético de las mismas.

Por tanto, puede pensarse que (intentar) poner límites desde el punto de vista penal al mal uso de estas tecnologías sobre los ciudadanos, hoy por hoy, dado que no se trata de riesgos actuales, sería una intervención meramente simbólica; sin eficacia real en el presente y orientada únicamente a establecer una línea político-criminal valorativa y declarativa hacia el futuro. Y puede ser así, como lo fue en su día la inclusión en el Código de los delitos de manipulación genética, por ejemplo. Pero, postergarla, podría hacer de ella una opción imposible cuando dicha tecnología esté socialmente implantada o se hayan generado daños irreparables, por ejemplo, para nuestra intimidad mental, antes de que se haya intervenido. El Derecho, también el Derecho Penal debiera acompañar al desarrollo de las nuevas tecnologías y, muy en particular, de las neurotecnologías.

3. ¿CÓMO PUEDEN INCIDIR LAS NEUROTECNOLOGÍAS EN EL COMPORTAMIENTO HUMANO? APUNTES PARA JURISTAS

3.1. Fundamentos (socio-)biológicos del comportamiento humano y neurotecnologías

A mi juicio, para poder valorar político-criminalmente lo que pudiera ser considerado un "*mal uso*" de una neurotecnología; para tratar de esbozar los límites del riesgo que resultaría aceptable en "usos socialmente admitidos y deseables" (*riesgo permitido*); o para fijar los ámbitos en los que tendrían incidencia, potencialmente lesiva o inocua, en primer lugar, hay que establecer en qué consisten, cómo se usan y qué efectos directos o indirectos pueden generar en el ser humano.

Pero para ello, a su vez, parece necesaria una breve exposición (orientado a lo que resulta relevante para un jurista) de los fundamentos (socio-)biológicos

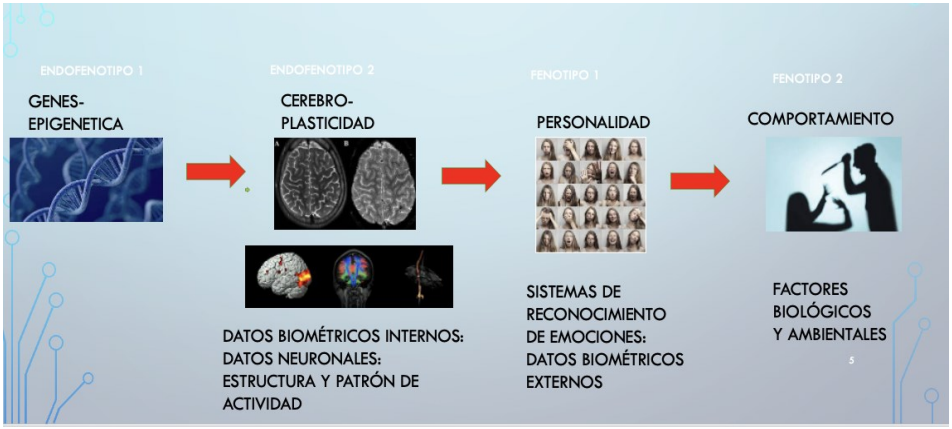
del comportamiento humano[2], porque, precisamente, las neurotecnologías tienen capacidad de incidir sobre nuestra actividad cerebral y, por tanto, sencillamente, pueden cambiar nuestro pensamiento, nuestra personalidad y nuestro comportamiento; pueden cambiar lo que somos, cómo somos y cómo nos comportamos. Para ilustrar esta idea, nos serviremos del modelo hipotizado por Adrian Raine en 2008 sobre las bases (socio-)neurobiológicas del comportamiento antisocial, según el cual habría un tránsito causal y co-explicativo que iría desde los genes (perfil genético) al cerebro (configuración estructural y funcional -actividad-), a una concreta personalidad fenotípica, que refleja lo anterior oculto (endofenotipo) y de ahí, al comportamiento antisocial. Obviamente, junto a nuestra biología y evolución, los factores ambientales juegan un papel co-protagonista para definir y para modificar todos esos niveles, pudiendo alterar, por ejemplo, la expresión final de los genes ante adversidades en la infancia (epigenética) o la configuración cerebral, p.e. con un daño cerebral adquirido o con la experiencia a lo largo de toda la vida, gracias a la *plasticidad* del cerebro… Es decir, este modelo no es un modelo determinista simple. Los factores ambientales, la experiencia a lo largo de toda la vida co-determinan lo que somos y cómo nos comportamos (Cfr., entre otros, magistralmente Sapolsky, 2018). Por ejemplo, con relación al genotipo del gen MAOA-L, interactuando con el maltrato severo o adversidad temprana en la infancia, aplicaríamos este modelo del siguiente modo (González-Tapia/Obsuth, 2015):

[2] Obviamente, lo que aquí se expone (con permiso y con petición de perdón incorporada a los neurocientíficos) no va a ser otra cosa que una descripción general y básica de algunas nociones que, desde mi perspectiva, un jurista debe conocer, aunque solo sea "en la esfera de lo profano". Además, dicho sea de paso, no es necesario argumentar la necesidad de aproximar el conocimiento neurocientífico y de la inteligencia artificial a los juristas y a los operadores jurídicos (y viceversa), encaminándonos a un mundo en el que dichas tecnologías disruptivas van a formar parte de nuestra cotidianeidad.

Ahora bien, dicho esquema no refleja suficientemente dos aspectos absolutamente centrales. En primer lugar, no refleja la enorme importancia que el sistema endocrino juega en la modulación de cerebro, (Sapolsky, 2004) y (Sapolsky, 2018)[3] y, de otra parte, lo que aquí más nos interesa, que este modelo explicativo del comportamiento humano no es un modelo estático, sino dinámico; es decir, que puede cambiarse a través de la incidencia, como se ha mencionado, de factores ambientales de todo orden. Así se refleja en esta otra figura, donde se refleja los potenciales cambios en la expresión de los genes y que puede transmitirse intergeracionalmente (epigenética) y la calidad moldeable del cerebro (plasticidad)[4].

[3] (Sapolsky, 2018) cit. *passim*. Este autor pone un ejemplo muy ilustrativo: un policía que ha sufrido una situación previa estrés relevante, tiene la amígdala activada (vía cortisol) y ello va a dar lugar a que, simplemente, vea más fácilmente una pistola que sacar un móvil en una situación compleja. Es decir, es más fácil que vea un estímulo amenazante (tendencia a ver amenazas) y a reaccionar como si fuera amenazante.

[4] En este esquema se muestran cuatro fases que irían desde la configuración genética que tenga el ser humano y que está oculto (por eso lo denominamos endofenotipo1) y que marcan la configurar del margen de posibilidades con las que, en principio cuenta una persona, en diversos aspectos. Por ejemplo, el gen el MAOA (Monoamina oxidasa alfa) que regula la serotonina en nuestro cerebro, se presenta en dos versiones básicas (L o baja actividad y H alta actividad). Pues bien, poseer la primera versión de este gen da lugar a una vulnerabilidad que, al interactuar con el ambiente (particularmente con el maltrato severo en la infancia) se produce una interacción que, vía sistema endocrino: estrés y cortisol) puede dar lugar a una afectación del "andamiaje socioafectivo" del cerebro, afectando a estructuras y al funcionamiento del cerebro. Y ello daría lugar a una específica configuración cerebral, tanto en lo que se refiere a la estructura como a la actividad cerebral (biomarcadores). Estos serían los datos neurobiológicos de la persona: datos biométricos internos: estructura y patrón de actividad neuronal de ese individuo, que también aparecen ocultos al exterior (endofenotipo 2). Esa configuración cerebral, unida por supuesto a la interacción permanente de factores ambientales, va a generar un patrón individual estable (pero también dinámico a través de la plasticidad del cerebro) que se manifiesta en la personalidad del sujeto, que recoge los rasgos del carácter de una persona: en este caso, daría lugar a un sujeto que, en principio tendría un carácter especialmente impulsivo y, conforme ese rasgo, expresará sus emociones al exterior (fenotipo 1). Por supuesto, el comportamiento humano es el fruto de un complejo conjunto de factores que, en esencia, pueden dividirse en ambientales (incluyendo la oportunidad) y también biológicos y entre ellos, estará esa marcada impulsividad, que lo hará simplemente más propenso a reaccionar violentamente ante cualquier estímulo amenazante real o considerado por el sujeto como amenazante en el momento de actuar.



Y ello, con relación a las neurotecnologías es particularmente importante, en la medida en que inciden, precisamente, en nuestro cerebro (endofenotipo 2), como un factor ambiental transformador directo y eficiente, *leyendo y re-escribiendo* nuestra actividad cerebral. Así, por ejemplo, detectando, aprendiendo y aprehendiendo "*atractores*" o patrones más estables de actividad neurológica, que pudieran reflejar, se imagina, un *pensamiento* o función cognitiva y transformando dicha información de entrada en otra de salida consistente en acciones a través de un robot, en lenguaje, visiones…Y en sentido contrario, generando información individualizada de salida, susceptible de transformar la original, cuando se usa, por ejemplo, en el tratamiento con implantes en supuestos de depresión severa o trastornos de estrés postraumático.

Por último, otro aspecto sin duda importante para un jurista con respecto a las neurotecnologías es entender cómo funciona el cerebro desde otra perspectiva y saber que el cerebro no solo reconoce y reacciona físicamente ante estímulos reales (ontológicos), sino también, y de la misma forma, ante una "realidad física pensada" y ante "realidad moral sentida" que activan al cerebro de la misma forma. Y es que, para comprender en toda su extensión la capacidad de las neurotecnologías para *re-escribir* (modular) nuestro cerebro, dado el carácter plástico de éste que lo permite, resulta muy ilustrativa la explicación que Robert Sapolsky da sobre el papel esencial que juega el córtex prefrontal en la toma de decisiones morales, pues es la parte del cerebro que nos permite tomar la decisión correcta, cuanto ésta es la opción más difícil en el caso concreto."[5] Y para ilustrar su funcionamiento, utiliza el ejemplo de la

[5] Sapolsky, Robert (2021). Neuroscience and the Law, *University of St. Thomas Journal of Law and Public Policy, 138*. Entre los factores biológicos que pueden incidir en el funcionamiento del córtex prefrontal y que pueden (co-)motivar decisión moralmente "injusta" o antijurídica, *(…)*

activación del córtex insular, mostrando que éste reacciona y manda señales a nuestro cuerpo (con participación del sistema endocrino) a partir de tres clases de estímulos: 1) la *experiencia* real del estímulo que sea, p.e. percibir algo que nos disgusta, como una hiena; 2) *pensar* sobre ese estímulo; y 3) *sentir* que algo, *moralmente* nos genera esa misma sensación de disgusto, dado que las emociones juegan un papel esencial en el juicio valorativo o moral. La reacción del cerebro y de nuestro cuerpo es la misma en esos tres escenarios. Por tanto, resulta importante pensar que las neurotecnologías pueden incidir en esas tres formas, haciendo experimentar cosas como reales, pensando sobre ellas al poderse incluir memorias, p.e.; o invocando emociones, que también pueden ser reguladas, inducidas o manipuladas. Y en este sentido, el papel que juegan las neurotecnologías es claro. La estimulación craneal profunda puede, directamente, activar el cerebro creando un estímulo artificial. Pero, más aún, incluso en experiencias de realidad virtual, en la medida en que nuestro cerebro reacciona también con la sola idea del estímulo, también podrá lograrse esa activación, porque el cerebro estará percibiendo un estímulo con un alto grado de realismo. O, más allá, en experiencias de los días u horas inmediatamente anteriores, por ejemplo, estar experimentando violencia o eventos estresantes a través de realidad virtual, tienen efectos de una cierta duración en el cerebro, que condiciona la respuesta que tengamos en una determinada situación. E igualmente, se pueden generar artificialmente asociaciones mentales condicionadas que generen sentimientos de afección u odio hacia cosas, situaciones, personas…

En definitiva, las neurotecnologías tienen la capacidad de modificar de forma directa y, sobre todo, de forma eficaz, nuestro cerebro; de modificar lo que pensamos, lo que nos gusta o nos miedo, de generar o modificar emociones, de cambiar nuestra personalidad y, en última instancia, de modificar nuestro comportamiento. El uso de neurotecnologías, invasivas o no, son un factor ambiental que tiene la capacidad de incidir de forma directa y de forma efectiva en nuestro comportamiento, en la medida en que tienen potencial para incidir y cambiar los anteriores escenarios: genes, cerebro y personalidad. Incluso, las menos invasivas, como la VR tienen ese potencial.

cuando la correcta es la más difícil, este autor señala los siguientes: 1) Inflamación neurológica de esa zona cerebral; 2) la concurrencia de estrés postraumático, que incrementa el volumen de la amígdala; 3) la presencia del parásito toxoplasma, que general impulsividad; 4) la existencia de un daño cerebral adquirido; 5) factores de adversidad severa en la infancia y adolescencia: p.e. pobreza, maltrato, en la medida en que puedan dar lugar a estrés crónico y a deficiencias en el PFC; 6) presencia de estrés prenatal que puede alterar la manifestación de los genes (epigenética); 7) el perfil genético del sujeto, p.e. MAOA-L + maltrato.

3.2. Efectos secundarios y riesgos descritos en el uso de las neurotecnologías sobre el ser humano

Sara Goering y colegas han alertado sobre la *espada de doble filo* que se cierne sobre las alteraciones neurológicas (neurointervenciones), hechas con interfaz cerebro-computadora, con estimulación craneal profunda o con cualquier otro dispositivo dirigido a incrementar capacidades cognitivas y motoras o a modificar o complementar dichas capacidades. La cuestión es que los impresionantes logros terapéuticos pueden venir acompañados de efectos que pueden reducir o confundir la conciencia sobre la propia identidad, individualidad o autonomía (agencia) (Goering/Brown *et al.*, 2021; Goering/Klein *et al.*, 2021 o Pugh, 2020 en cuanto a los términos del debate.).

Con respecto a la estimulación craneal profunda (DBS, en sus siglas en inglés) la literatura científica ha descrito dos tipos de riesgos o efectos secundarios. Como efectos secundarios menores, se ha descrito hormigueo pasajero y problemas en la modulación del lenguaje, en pacientes tratados con desórdenes motores, como el Parkinson. Sin embargo, algunos usuarios también han experimentado efectos secundarios más importantes, como un incremento significativo de la impulsividad, adicción al juego o agresividad, manía, sensación de sentirse robótico o con inseguridad sobre la autenticidad o autoría de sus sentimientos o comportamientos; o tener la sensación de que la máquina decide por ti, en dispositivos diseñados para controlar deseos y tratar trastornos como la anorexia, depresión y obesidad. También se han descrito cambios de valores y comportamiento. Por ejemplo, pacientes de Parkinson tratados con DBS han experimentado hipersexualidad y otros apatía o sentimientos de no reconocerse a sí mismos. Y ello, se afirma, pudiera incrementarse de forma radical en tratamientos que impliquen, por ejemplo, recuperación o reestructuración de memorias perdidas o supresión de éstas, como ocurre en tratamientos relativos al Alzheimer o al trastorno por estrés postraumático (PTSD). En algunos usuarios con problemas mentales, como depresión mayor o TOC, se han descrito problemas de alienación (no sentirse ellos mismos) - (Cfr. Goering *et al.*, 2017).

En cuanto a los usuarios de dispositivos interfaz cerebro-computadora (BCIs), se ha descrito igualmente un alto nivel de satisfacción con el dispositivo pero efectos secundarios similares, particularmente en lo referido a una menor sensación de control sobre el dispositivo inteligente, sintiendo que actúa de forma independiente o como algo así como el funcionamiento de un "*autocorrector*", es decir, sentir que la máquina es quien interpreta y define completamente la orden que el usuario *estaba intentando* dar y la ejecuta, cuando

se siente que, tal vez no era eso exactamente lo que el sujeto quería expresar[6]. Se han transmitido expresiones de los usuarios, tales como: "You just wonder how much is you anymore," said one. "How much of it is my thought pattern? How would I deal with this if I didn't have the stimulation system? You kind of feel artificial.", expresando problemas de agencia y sentido de la propia identidad como agente que actúa conforma a sus propias opciones y valores (Klein *et al.*, 2016). Y es que con la inclusión de la inteligencia artificial en el procesamiento e interpretación de las señales cerebrales ha permitido un avance espectacular, pero ello trae consigo que la interpretación de la información cerebral la hace un algoritmo (en gran medida opaco), que funciona esencialmente a base de predicciones. Ello introduce una fase en el proceso que realmente se desconoce entre los pensamientos del usuario y la información de salida que vuelve a él tras ser captada por un dispositivo, tratada con un algoritmo y remitida la información de salida hacia el cerebro del usuario. Y dicha información de salida puede ser una correspondencia real con el pensamiento del sujeto o puede ser una predicción del algoritmo de lo que el sujeto quiere, de lo que va a hacer a continuación… Como un *autocorrector*, facilita las cosas, pero muchas veces decide por ti porque es más eficiente; o lo hacemos con una especie de agencia híbrida…(Cfr. Drew, 2019)

Otros pacientes han manifestado sentir una "simbiosis radical" o experimentar cambios de forma de ser o pensamiento, como un usuario que cayó en la ludopatía y que solo pudo comprender la situación en la que se hallaba y la problemática generada en su entorno cuando fue "desconectado" o cuando, a consecuencia de la desaparición de la empresa suministradora, hubo de retirarse perdiendo su "nuevo yo" (Gilbert *et al.*, 2019). Esta alteración del sentido propio de agencia, además de los problemas de identidad personal (alienación), puede plantear otros problemas jurídicos relativos a la propia consideración de agente (pertenencia o imputación de la acción al sujeto) y, por tanto, a su responsabilidad y ambos pueden ser todavía más complejos en los supuestos de interfaz cerebro-cerebro (B2B, BBIs) por cuanto dos personas intervienen en el proceso y se generaría una especie de "control-compartido" y pertenencia del comportamiento y una "identidad híbrida" o mixta, con transferencia de memorias, entre persona y animal…(Cfr. Hildt, 2015)[7].

[6] (Kögel *et al.*, 2020) Este estudio desarrollado con usuarios de BCIs con problemas motores graves, destacan la importancia de mantener el sentido de autonomía y de agencia en la operatividad de estos dispositivos y la dificultad que entraña mantenerse "libre de emociones" para lograr un mejor entendimiento con el dispositivo. En cualquier caso, la conclusión general que se extrae de estos estudios es que, en general, los usuarios manifiestan un alto un nivel de satisfacción.

[7] También se han puesto sobre la mesa problemas técnicos y operativos nada desdeñables. Por ejemplo, Gabriel Silva, 2018, apuntaba a la necesidad de *individualizar* estos dispositivos para

(…)

Con respecto a los tratamientos terapéuticos de los problemas mentales, el primer problema añadido que surge es, precisamente, que lo que se busca directamente con el tratamiento es cambiar la mente del paciente p.e. en la anorexia (Drew, 2019), en el sentido de que generalmente se busca leer e influenciar las emociones y los estados de ánimo del paciente. Con respecto a las denominadas "*Afective BCIs*", Steiner y Friedrich apuntan a la especial afectación que el uso de estas tecnologías puede tener en la identidad y en la autonomía del ser humano, en la especial capacidad para influir y manipular a individuos y grupos, por el papel crucial que la emoción juega en la toma de decisiones, "moldeando intenciones, decisiones y acciones" (Steinert/Friedrich, 2020).

Como ha afirmado Rafael Yuste, la tecnología implantada es mucho más potente, pero por su propia naturaleza siempre estará en el ámbito de la salud o, siquiera, bajo el control médico; sin embargo, los mayores problemas éticos y jurídicos provienen de la no implantada o no invasiva, puesto que se hará de ella un uso comercial (producto electrónico de consumo) y, por tanto, accesible en masa a toda la población y no está regulada (Ansede, 2022).

4. PROTECCIÓN PENAL DE LOS NEURODERECHOS: PROPUESTA POLÍTICO-CRIMINAL

Ante el avance conjunto e interconectado de las neurociencias, la inteligencia artificial y la robótica y ante los riesgos se están apuntado… ¿debe el Derecho, y el Derecho Penal en concreto, entrar a regular los lineamientos básicos entre los que debe discurrir el avance de estas tecnologías? La respuesta parece no suscitar dudas: sí, dada la relevancia constitucional de los bienes jurídicos que pueden verse afectados. Teniendo en cuenta los riesgos que se han

cada usuario en orden a su rendimiento óptimo y a las dificultades de compatibilizarlo con un desarrollo mayor y una implantación generalizada entre la población. ¿Cómo y dónde almacenar en la nube toda la información generada? Siendo realistas… probablemente se trate de dispositivos que, para determinadas áreas, p.e. motora, se conformen con un conocimiento de base y procedimiento de aprendizaje altamente, por así decirlo, de serie; que será más significativo cuanto más amplio sea su "potencial mercado" de usuarios consumidores…. *¿La fabricación a gran escala que se presume para hacer todo más eficiente y rentabilizar costes, es compatible con la individualización y con la privacidad?* Probablemente ello sólo será viable para minorías con alto poder adquisitivo. Para el conjunto de la sociedad, cualquier tecnología que se generalice (sobre todo de forma inhalámbrica) habrá de tener un inevitable diseño "en serie", más o menos perfeccionado y con un mínimo de adaptabilidad al usuario. Y en tal escenario, la rentabilidad provendrá, probablemente, del contenido que se proporcione a través de tales dispositivos; más allá de uso terapéutico, por ejemplo, para manipulación de la opinión pública, para acallar movimientos disidentes del pensamiento hegemónico, para vender productos… Y otra cuestión, *¿se pueden retirar estos dispositivos cuando, por ejemplo, el usuario no los puede continuar pagando o no los puede reparar? ¿Y si desaparece la empresa que te lo ha implantado o lo mantiene? ¿Pueden este tipo de dispositivos quedar sometidos a las leyes del capitalismo?*

descrito, los derechos individuales fundamentales que pueden verse afectados serían, al menos, la salud y la integridad física y mental, la libertad y autonomía individual, la dignidad o integridad moral y la intimidad.

Se ha indicado también que hay dos riesgos centrales. El primero es general y radical, que tiene que ver con la transformación de la naturaleza humana hacia su hibridación, a través de las neurotecnologías, mediante de la interconexión simbiótica del cerebro humano con un dispositivo externo de computación. A mi juicio, a la vista de los beneficios y extraordinarios avances que permite en el campo de la salud, el uso terapéutico no puede ser rechazado. No parece razonable ni realista prohibir tales prácticas, cuando se haga conforme a la *lex artis*, pues contradirían el art. 9.3. de la CE, que obliga a remover los obstáculos que pudieran tener los ciudadanos para hacer efectivos sus derechos y su participación en la sociedad. En este ámbito, por tanto, las garantías deben provenir del consentimiento informado y del respeto al código bioético vigente. Ahora bien, fuera del uso terapéutico, ¿es admisible el uso de neurotecnologías invasivas o implantadas? ¿sería admisible o se justificaría un uso militar? ¿podría justificarlo la seguridad pública, por ejemplo, para el control de delincuentes especialmente peligrosos? ¿puede ofrecerse a los consumidores a demanda p.e. para mejorar sus capacidades motoras y/o cognitivas? ¿se pueden utilizar subliminalmente para manipular a individuos y grupos? ¿otros medios menos eficaces y no invasivos (diademas, gorras o cascos…) serían admisibles en otros usos no terapéuticos, p.e. realidad virtual para aprendizaje o para usos lúdicos?

Ninguna de esas preguntas, tampoco la inicial, puede contestarse desde una posición ideológicamente neutral, suponiendo que tal cosa realmente pudiera darse. Contestar a la pregunta de si un delincuente (o potencial delincuente) debería recibir un tratamiento neurológico, en caso de estar disponible, voluntario o no, reversible o no, con fármacos, cirugía o dispositivos, que "borrara" su inclinación hacia la violencia o hacia la pederastia… no puede contestarse en términos categóricos; ni puede tampoco contestarse desde la equidistancia. En absoluto. Requiere una respuesta valorativa, ideológicamente conformada. El Derecho, por supuesto tampoco el Derecho Penal no es ideológicamente neutral; no lo puede ser, porque no lo admite la propia noción de normatividad, indefectiblemente unida a la idea de valoración; ni siquiera desde planteamientos estrictamente funcionalistas. Y, como expresara magistralmente Elías Díaz, en nuestra Constitución los pilares ideológicos fundamentales son el art. 10 y el art. 9, pues entre ambos sintetizan la fórmula concreta de Estado Social y Democrático de Derecho que consagra nuestra constitución. Y en última instancia, en caso de conflicto, el diseño constitucional se escoraría hacia una posición liberal y humanista de los derechos fundamentales, con el ser humano individual como centro del sistema

y no el grupo (Cfr. M.I. González-Tapia, 2013). Desde esta perspectiva, parece claro que la política-criminal en materia de derechos fundamentales debe estar orientada a la maximización de la protección de los derechos fundamentales de los ciudadanos frente al Estado y frente al mercado, con el límite principal en el respeto de los mismos derechos para los demás.

Dentro de esta perspectiva liberal y humanista, se encuadrarían propuestas como la del *Grupo Morningside*, liderado por Rafael Yuste, para quienes la interacción neurociencia e inteligencia artificial requiere incorporar la tutela jurídica de una serie de nuevos derechos, denominados neuroderechos, que están dirigidos a la protección de la integridad y la indemnidad mentales. Así mismo, con buen criterio, reclaman que dicha protección se haga a nivel internacional, como la que se dispensa en la Declaración Universal de los Derechos Humanos de 1948 y otras declaraciones análogas posteriores[8]. Este planteamiento ha sido acogido en Chile en una muy meritoria propuesta de ley de desarrollo de los derechos fundamentales reconocidos en la Constitución chilena y que todavía está en tramitación. De la mano de ambos, iremos desgranando la propuesta que formulan, relativa a los protección internacional y constitucional de los neuroderechos (*neuroprotección*), que asumimos en su integridad y que, simplemente, enfocamos más específicamente hacia el Derecho Penal. Y ello, porque una protección internacional y constitucional de estos neuroderechos lleva de la mano su protección penal, de acuerdo con los principios limitadores del bien jurídico, de intervención mínima y de proporcionalidad.

4.1. Protección de la salud e integridad mental: principio de precaución en el uso de las neurotecnologías

En una sociedad democrática, el uso de las neurotecnologías debe guiarse por el objetivo del "bien común" y "al servicio del ser humano", por encima del lucro y del control social. En correspondencia con ello, ante riesgos difusos con incertidumbre acerca de su envergadura y de su reversibilidad, debe atenderse al principio de precaución que proteja la salud mental individual y pública de la ciudadanía. En base a ello, se propone la restricción de las neurotecnologías de eficacia directa sobre el ser humano al uso terapéutico y la implantación

[8] Cfr. para todo lo que sigue, en cuanto a la propuesta del grupo: (Yuste *et al.*, 2017) y (Goering/Klein *et al.*, 2021). En cuanto a los neuroderechos, sobre el concepto y evolución de estos nuevos derechos: (Cfr. Ienca, 2021). En cuanto al Proyecto de Reforma constitucional en Chile, pueden consultarse los documentos y los trabajos parlamentarios en: https://www.camara.cl/legislacion/ProyectosDeLey/tramitacion.aspx?prmID=14385&prm BOLETIN=13828-19; y https://www.diarioconstitucional.cl/articulos/proyecto-de-ley-sobre-neuroderechos/

limitada, particularmente con respecto a sus contenidos y grado de exposición, para otras neurotecnologías como la realidad virtual.

El uso de las neurotecnologías, como se ha dicho, debe guiarse por el principio de precaución a la hora de definir el riesgo permitido y situar el eje central en los potenciales daños que puedan derivarse para los derechos fundamentales del ser humano. Desde esta perspectiva, y en atención a los conocimientos actuales, parece obligado restringir el uso de las neurotecnologías más eficaces, invasivas o no, de incidencia directa sobre el ser humano, a un uso terapéutico, sujetas, por tanto, al control ético de este ámbito. Hoy por hoy, el principio de precaución debe priorizarse en el ámbito de las *neurointervenciones*, puesto que en torno a ellas todavía *no hay un conocimiento científico asentado*, existiendo incertidumbre en torno a las técnicas que están en pleno desarrollo, a los *efectos* que, a corto, medio o largo plazo pudiera generar su uso; a las *transformaciones* y efectos colaterales que, en virtud de la plasticidad del cerebro, pudieran generarse en su salud o en su personalidad y en el carácter *reversible* o no de los efectos de estos tratamientos… Y lo que está en juego es la salud mental (y física) de la persona, en cuyo ámbito habrán de estar incluidos los atentados lesivos contra la salud e integridad mental del individuo, entre los delitos de lesiones y su tratamiento jurídico, en paralelo al que actualmente se dispensa a los tratamientos médicos curativos (terapéuticos) o no curativos, sin perjuicio de su proyección también en torno a la salud pública.

Frente a neurotecnologías menos eficaces y no invasivas, la cuestión es mucho más compleja, pues aparecen más difusos y desconocidos los riesgos a la identidad y a la autonomía que pudieran generarse en los usuarios, p.e., fruto de manipulaciones emocionales dirigidas a orientar el voto en unas elecciones, a comprar algunos productos, a crear estados de opinión frente a algo o alguien…[9] Además, la reflexión jurídica se ve empañada y seriamente dificultada por la carrera acelerada y paralela de las compañías para su implantación comercial y en masa en el conjunto de la sociedad. Probablemente, el conato de regulación llegue cuando, de hecho, esté implantada ya en la sociedad y se haya hecho sin control normativo suficiente. A mi juicio, el nivel de protección deberá formularse en un juicio valorativo que priorice la salud e integridad mental de los ciudadanos, y su libertad frente a manipulaciones, en base igualmente al principio de precaución. Deberá definirse el riesgo permitido a partir de la clarificación previa de los riesgos para la salud física mental del usuario por parte de la ciencia (sobre lo que no hay hasta hoy un

[9] Ello se recoge como práctica prohibida en el art. 5 de la Propuesta de Reglamento de Ley de Inteligencia Artificial, apartados a) y b).

cuerpo doctrinal relevante[10]) y, a partir de ahí, definir el riesgo permitido sobre la base de usos socialmente adecuados, de clara utilidad y beneficio para el conjunto de la sociedad, frente usos abusivos efectivos o razonablemente previsibles que de ellas pueda hacerse. El ejemplo por antonomasia de la problemática de este tipo de neurotecnología es la realidad virtual, que está siendo implantada masivamente en la sociedad, con proyectos como Metaverso que van a ahondar definitivamente en esta implantación, sin que ni siquiera se sea todavía consciente de los efectos que ello pudiera generar en la salud pública de los ciudadanos, ni exista un debate jurídico serio y ajeno a las multinacionales interesadas en torno, por ejemplo, a sus contenidos.

4.2. Tutela de los datos neurológicos individuales como órgano: privacidad (e indemnidad) mental

En la intersección entre la neurociencia y la inteligencia artificial, se ha dicho que el primer (neuro)derecho que debe tener el ciudadano es el de tener y mantener sus datos neurobiológicos en la privacidad, reservando para ellos un protección reforzada, en una doble vertiente: otorgándoles el nivel de confidencialidad de los datos de salud especialmente sensibles, como el que se dispensa a los órganos o tejidos humanos; y atribuyéndoles la condición de órgano, en orden a controlar y limitar su tráfico y a excluir su explotación con ánimo de lucro.

En cuanto a la protección de este derecho o bien jurídico individual, debe formularse una protección efectiva ante la *captación directa* de datos neurológicos del usuario, identificado o identificable, p.e. estableciéndose la opción de *no compartir esta clase de datos por defecto* en los dispositivos, la necesidad de renovar ese consentimiento para seguir compartiendo y establecer un cuidado y claro *consentimiento informado* que realmente sea informativo para el conjunto de la ciudadanía[11], donde se explique claramente los propósitos para

[10] De la realidad virtual se sabe a través de una vasta literatura, sobre todo, la eficacia muy significativa que tiene para el aprendizaje de conocimientos y destrezas, por ejemplo. También se está haciendo un uso cada vez más intensivo de ella en el tratamiento de la rehabilitación física y de la salud mental, por ejemplo, entre otros muchos (González Moraga *et al*., 2022) (Monaghesh *et al*., 2022) (Mazza *et al*., 2021)… También se sabe que Meta, Microsoft y otras compañías están invirtiendo cantidades ingentes de dinero en sus proyectos lúdicos o de interacción social global en un contexto de realidad virtual: p.e. Metaverso… (Maloney, 2021) Pero… ¿qué se sabe realmente de los riesgos? Es un problema que, hasta donde alcanzo a conocer, está todavía en un estudio inicial y, probablemente, la idea más aproximada que os podemos hacer de los mismos es, precisamente, a través de su eficacia en los usos que se están haciendo y en algunos trabajos comienzan a explorar la dimensión ética y los riesgos de la realidad virtual (por ejemplo, el ilustrativo e intersante trabajo de (Rueda / Lara, 2020) o (Kaimara *et al*., 2021).

[11] En la exposición del *Proyecto de ley, iniciado en moción de los Honorables Senadores señor Girardi, señora Goic, y señores Chahuán, Coloma y De Urresti, sobre protección de los neuroderechos y la (…)*

los que se van a ceder sus datos y por cuanto tiempo… Junto a ello, se deben diseñar procesos de captación, almacenaje y tratamiento de la información que salvaguarden la privacidad y el rastreo y la auditoría del procesamiento y tráfico de la información, p.e. utilizando tecnología *blockchain* o con códigos abiertos y transparentes. Así mismo, debe protegerse al ciudadano frente a la *captación indirecta* de tales datos de los usuarios que no comparten sus datos. Teniendo en cuenta la capacidad predictiva de la inteligencia artificial, se deberá proteger a estos ciudadanos de una extracción indirecta de sus datos neuronales a través de la restricción y férreo control del tráfico de datos neurológicos, tal y como se hace para otros órganos humanos, que son *res extracommercium*. Y ello, porque (recuérdese las figuras propuestas arriba) se puede extraer información neurológica (endofenotipo) a través de patrones de comportamiento, emociones, lenguaje y otros datos biométricos obtenidos a través de otros dispositivos (fenotipo y comportamiento exteriorizado) e, incluso, ser completada a través de la estadística, de los patrones predictivos extraídos a través de usuarios que sí comparten sus datos (Vid. así Ienca, 2021).

Para entender la repercusión de lo que estamos hablando, basta señalar por ejemplo la detección temprana de enfermedades que puede hacerse, entre otros medios, a través del patrón de tecleo en el móvil con relación al Párkinson, de deambulación con respecto al Alzheimer o patrones de atención… Así mismo, datos relativos a navegación por las redes sociales y el análisis de los mensajes arrojados en ellas, permiten detectar riesgos como comportamientos o ideaciones suicidas; datos obtenidos en el entorno laboral sobre las emociones sentidas por sus trabajadores, pueden informar sobre su motivación… No hace falta explicar… que tales datos son más que interesantes para vender productos, establecer primas de seguros, seleccionar potenciales parejas, en la selección de candidatos a un puesto de trabajo…

Probablemente, la protección de la privacidad en este ámbito requeriría formular una protección específica y reforzada dentro de los delitos contra la intimidad y reformar también el delito de tráfico de órganos para incluir

integridad mental, y el desarrollo de la investigación y las neurotecnologías (Boletín nº 13.828-19), en adelante Proyecto-Chile, se sitúa el eje central de la neuroprotección en el consentimiento informado, estableciéndose "la prohibición de cualquier forma de intervención de conexiones neuronales o cualquier forma de intrusión a nivel cerebral mediante el uso de neurotecnología, interfaz cerebro computadora o cualquier otro sistema o dispositivo, sin contar con el consentimiento libre, expreso e informado, de la persona o usuario del dispositivo, inclusive en circunstancias médicas. Aun cuando la neurotecnología posea la capacidad de intervenir en ausencia de la conciencia misma de la persona." (art. 3). Así mismo, en el art. 5 se precisa que habrá de informarse al usuario de los posibles efectos físicos de su aplicación, los eventuales efectos cognitivos y emocionales de los mismos, los derechos y deberes, normas sobre privacidad y protección de la información, medidas de seguridad adoptadas, y contraindicaciones.

transferencias no autorizadas de neurodatos individuales, con ánimo de lucro o comercial, en el bien entendido de que, al tratarse de bienes jurídicos individuales, habrían de circunscribirse a captación/interceptación, almacenamiento, procesamiento y/o tráfico de datos de personas identificadas o identificables.

4.3. Tutela penal de la identidad personal y la autonomía ante las neurotecnologías

El tercer pilar de la neuroprotección habría de estar situado en la tutela de la identidad y el sentido de agencia o autonomía personal[12]. La identidad personal, en tanto que integridad corporal y mental, que se identifica con el sentido del propio yo, que nos permite reconocernos como nosotros mismos en nuestra individualidad y en el contexto espacial y temporal en el que nos hallamos, incluyendo nuestra narrativa personal, nuestra memoria, como nos vemos y reconocemos, lo que amamos, nuestra personalidad. Ienca (2021, p. 5 y 6) dice que se trata de un derecho que presupone la compleja noción de personalidad y que habitualmente se define como el haz o conjunto de propiedades que definen a alguien como persona singular o hace de alguien la persona que es, y que la distingue de todos los demás. Y en el ámbito normativo, el derecho de cada persona a formar su propia identidad y consciencia (libre desarrollo de la personalidad), sin injerencias ilícitas externas. En cuanto a la noción de agencia, equivaldría a la noción de autonomía personal, a la libertad de pensamiento, en tanto que sujeto capaz de regirse por sí mismo, de tomar sus propias elecciones y de quien se predica capacidad de acción (capacidad para realizar un comportamiento exteriorizado, activo u omisivo, de naturaleza voluntaria, es decir, que le pertenece como agente y así se le imputa o atribuye por parte del Derecho). Yuste *et al*., (2021) la han definido como la libertad de pensamiento (o conciencia) y la libertad de elección de las propias acciones ("*freedom of thought and free will to choose one's own actions*"). Se ha dicho de este derecho que es la matriz de todas las libertades, un principio axiomático sin el que no pueden entenderse ni existir todas las demás libertades del ser humano (Cfr. Ienca, 2021) y su engarce

[12] En el Proyecto de Chile esta protección se prevé en el art. 4, donde se propone lo siguiente: "Artículo 4: Queda prohibido cualquier sistema o dispositivo, ya sea de neurotecnología, interfaz cerebro computadora u otro, cuya finalidad sea acceder o manipular la actividad neuronal, de forma invasiva o no invasiva, si puede dañar la continuidad psicológica y psíquica de la persona, es decir su identidad individual, o si disminuya o daña la autonomía de su voluntad o capacidad de toma de decisión en libertad. - El límite de cualquier intervención de conexiones neuronales será siempre la protección de los sustratos mentales de la identidad personal. - Las únicas excepciones admitidas a la alteración de la continuidad psíquica o autónoma serán en casos de investigación o terapia clínico-médicas, en cuyo caso se aplicará el código sanitario vigente"

constitucional se ubica en el artículo 10 de la Constitución, donde se consagra a la dignidad del ser humano y al libre desarrollo de su personalidad, como ejes de todo el entramado de derechos fundamentales.

A la vista de los efectos secundarios y riesgos reportados de la aplicación clínica de las neurotecnologías, obviamente, es preceptivo el consentimiento informado del paciente, con expresa instrucción sobre potenciales cambios en la identidad, en personalidad, en el estado de ánimo y en el comportamiento del usuario, el carácter reversible o no de dichos cambios, de acuerdo con el conocimiento actual; ser instruido en cuanto a las opciones y consecuencias de interrumpir el tratamiento.... En un modelo pluriofensivo, comprensivo de una tutela indirecta de la salud mental y de la libertad y de la dignidad del ser humano, los atentados a estos derechos deberían articularse sistemáticamente a través de la protección de la integridad moral. Así, sin perjuicio de las relaciones concursales que pudieran establecerse con respecto a daños concretos en cuanto a la salud mental (que pudieran probarse) o acciones dirigidas o manipuladas a las que el sujeto hubiera sido determinado o conducido, las neurointervenciones (invasivas o no) no consentidas que hayan anulado la libertad del individuo o lo hayan manipulado subliminalmente a través de la inteligencia artificial, deben ser articuladas como atentados contra la dignidad e integridad moral del ser humano.

4.4. Tutela de la igualdad y la no discriminación con respecto a las neurotecnologías

La problemática derivada del respeto al derecho a la igualdad en cuanto a las neurotecnologías y a la inteligencia artificial se plantea en dos ámbitos esenciales. De una parte, la protección de la igualdad frente a los *sesgos* discriminatorios que articularían y dirigirían la decisión algorítmica; y, de otra, los problemas o riesgos previsibles que se anticipan en conexión con las posibilidades de mejora o expansión de las capacidades cerebrales a través de las neurotecnologías, invasivas y no invasivas, y en cuanto a su uso terapéutico o comercial.

Así, dejando aparte el problema de los sesgos en la inteligencia artificial, otro de los objetivos centrales de la neuroprotección sería la tutela de la igualdad frente al uso discriminatorio de las neurotecnologías y también la salvaguarda del acceso *equitativo* a las mismas por parte de la población. El tratamiento jurídico de este problema, a mi juicio, requiere diferenciar el *uso terapéutico o curativo*, dirigido a paliar o mejorar las condiciones de vida de personas que sufren discapacidad física o mental, p.e. que permitan a una persona ciega la visión, a una persona tetrapléjica poder desplazarse, a una persona poder hablar, a curar una enfermedad mental; y el tratamiento *no curativo sino meramente expansivo* de las mismas, de amejoramiento físico,

cognitivo, mental o emocional de los seres humanos, p.e. dispositivos que nos conecten con una computadora y que nos abran al conocimiento y la comprensión del mundo, en todos los campos del saber, en todos los idiomas…; o dispositivos que mejoren nuestro rendimiento motor, o nuestra visión o nuestro oído…; o que mejoren o amplifiquen nuestra empatía… a través de una interfaz cerebro-computadora…

Este problema del "mejoramiento cerebral" es, a mi juicio, uno de los más complejos desde el punto de vista jurídico, porque sobre la base del carácter ("*indiscutiblemente*") positivo del objetivo final, se esconden otros problemas secundarios derivados del impacto real que tendría para el conjunto de la sociedad la implantación social de estas técnicas, particularmente en usos no curativos. Y es más difícil, porque la problemática que suscita es de justicia social, de equidad, de la visión acerca del progreso social que nos proporciona… y todos ellos son nociones difíciles de concretar en el ámbito jurídico, sin olvidar que debe pensarse en el contexto internacional en el que va a desarrollarse. Este es un ámbito todavía más complejo desde el punto de vista jurídico, porque tiene contornos imprecisos como la justicia y la equidad y plantea problemas valorativos "en los márgenes" y en la "zona gris", tan difícil de delimitar para el Derecho Penal, a la hora de identificar bienes jurídicos susceptibles de protección y el nivel de protección que debe proporcionar el Derecho Penal, atendiendo al principio de intervención mínima. Piénsese en un primer problema a la hora de identificar si socialmente la hibridación de la especie humana seria ampliamente admitida o no; de quienes creen que debe darse un uso meramente terapéutico y quienes puedan entender que debe primar la libertad de quien quiera (y pueda costear) ser mejorado, como por ejemplo, una operación de cirugía estética; en problemas de discriminación entre personas mejoradas y no mejoradas, como una nueva brecha social irreversible como ha afirmado Rafael Yuste; de si debe hacerse un acceso prioritario (discriminación positiva) de estas tecnologías como instrumento para proteger y mejorar a las personas que presenten algún tipo de discapacidad o no; de si debe abrirse el acceso a cualquiera y no establecer ninguna protección para las personas no mejoradas frente, por ejemplo, un empleo público, unas oposiciones, una competición deportiva…¿Quién va a ser el referente del ser humano en nuestro futuro, sobre el que debe articularse prioritariamente la protección o interés jurídicos? ¿el natural o el híbrido?

Humildemente, no podría contestar a estas preguntas. Con la misma humildad, creo que, hoy por hoy, jurídicamente nuestro referente debe seguir siendo el ser humano *natural* y la justicia social reflejada en el art. 9.3 de nuestra Constitución (en tanto no haya un acceso universal o, cuando menos, mayoritario, y libre a estas posibilidades de mejora) parece imponer una priorización del uso terapéutico con fondos públicos de estas neurotecnologías;

de agudizar el consentimiento informado y el principio de precaución en usos no curativos; y también el establecimiento de garantías de no discriminación entre los ciudadanos y, muy en particular, en el ámbito de lo publico: por ejemplo, en el acceso a la función pública, competiciones oficiales... Y la protección penal habrá de articularse, a mi juicio, a través de la protección de bienes jurídicos supraindividuales sobre los que se materialicen contravenciones graves de estos principios de igualdad material y justicia social.

Así mismo, con su fundamento en la igualdad, en la dignidad y en el libre desarrollo de la personalidad, se plantea el problema añadido de las disparidades culturales que existen a nivel internacional, por ejemplo, con respecto al valor del ser humano individual frente al grupo, al nivel de protección e importancia de la privacidad y la identidad… Ello, no solo hace muchísimo más complejo (unido a los inimaginables intereses lucrativos que subyacen en todo este negocio) establecer un catálogo de neuroderechos, con un mínimo esencial uniforme de protección universal. También hace mucho más difícil al ciudadano que quisiera proteger su individualidad de "persona limitada y (im)perfecto" y con el máximo valor por sí mismo en su individualidad y que, simplemente, no quiera "mejorarse", ni uniformizarse … Cómo resistirse a la presión cuando, por ejemplo, ello te impida acceder a determinados puestos de trabajo, tenga repercusiones sociales negativas… Cómo va a poner límites un Estado … y a través del Derecho Penal... cuando otros países no lo hagan… Cómo proteger también, en definitiva, el *libre* acceso a dichas neurotecnologías.

5. CONCLUSIÓN

A mi juicio, queda patente que la reflexión en torno a los riesgos derivados del uso de las neurotecnologías y de los potenciales lineamientos político-criminales de la protección de los denominados neuroderechos es urgente; y es necesaria. En este trabajo se aportan, humildemente, algunas ideas; una primera aproximación en cuanto a los dominios y los criterios que habrían de presidir la intervención penal en este ámbito. Para ello, en primer lugar se ha analizado el estado del arte de la neurociencia en cuanto a la potencial incidencia de las neurotecnologías en el comportamiento humano. Según los fundamentos socio-biológicos del comportamiento humano, se pone de manifiesto que las neurotecnologías tienen la capacidad de modificar de forma directa nuestro cerebro, nuestra personalidad y nuestro comportamiento; y se exponen los principales riesgos o efectos segundarios nocivos descritos hasta hoy por la literatura, destacando que el principal problema es el del *doble uso* que puede tener esta tecnología. En atención a tal potencial y a tales peligros, el objetivo de este trabajo ha sido el de formular una revisión

político-criminal (siquiera los rudimentos iniciales) sobre la base de su necesidad ante práctica ausencia de estudios penales específicos sobre esta materia. Sobre la base de la propuesta de Yuste *et al.* (2017 y 2021), descendiendo del ámbito del reconocimiento internacional o constitucional de los neuroderechos al de su tutela penal.

Sobre la base de todo lo anterior, se propone, en esencia, lo siguiente: 1) Con respecto a la integridad mental, establecer precisiones en el ámbito de los delitos de lesiones y restringir el uso directo de las neurotecnologías más eficaces, como la estimulación craneal profunda y o la interfaz cerebro-computadora, invasivas o no, a las (neuro-)intervenciones consentidas, en el ámbito médico y con un uso terapéutico, en base a las incertidumbre todavía existente en cuanto a técnicas y riesgos y, por tanto, al principio de precaución. 2) En el ámbito de la privacidad, se aboga por una protección reforzada de la privacidad de los datos neurológicos, de persona identificada o identificable, en el ámbito de los delitos contra la intimidad y su consideración como un órgano del ser humano *(res extracommercium)*, otorgándole la misma protección que a los demás órganos, lo que requeriría precisiones en torno al delito de tráfico de órganos. 3) En cuanto a la protección de la identidad personal y la autonomía o agencia, se deriva su protección a los delitos contra la integridad moral, tanto en intervenciones directas como subliminales. 4) En cuanto a la igualdad, se asume la necesidad de acoger la tutela de la igualdad frente al uso discriminatorio de las neurotecnologías y también la salvaguarda del acceso *equitativo* y *libre* a las mismas, proponiendo que la intervención penal se articule a través de la protección de bienes jurídicos supraindividuales sobre los que se materialicen contravenciones graves de estos principios de igualdad material y justicia social, particularmente en el ámbito de *lo público*. Se aboga aquí por la necesidad de intervención del derecho penal en todos estos ámbitos para establecer límites, que pivotan sobre una idea liberal y humanista del ser humano. Quedan abiertas dos áreas, en los márgenes, particularmente complejas, que son el amejoramiento cerebral con fines no terapéuticos y el uso de la realidad virtual.

Finalmente, dicho todo lo anterior… no se ocultan las serias dudas que se albergan sobre el papel que el Derecho Penal está llamado a jugar en el contexto actual; menos aún, en el devenir futuro que se prevé de interacción entre neurociencia e inteligencia artificial, en todos los órdenes de la vida…. Resulta fácil imaginar que tal cambio de contexto socio-económico, pueda generar un nuevo paradigma valorativo respecto del que el Derecho Penal actual requiera cambios sustanciales; una política-criminal sustancialmente distinta. No se ocultan las serias dudas sobre si el Derecho Penal tiene todavía capacidad realmente para ejercer una protección efectiva frente a estas nuevas formas de dañar a bienes jurídicos centrales y radicales, cuando los gobiernos más

poderosos y las empresas que gobiernan nuestro mundo están invirtiendo ingentes cantidades de dinero en el desarrollo de estas tecnologías y su objetivo es la implantación general y su uso masivo en todos los órdenes de la vida… No se ocultan serias dudas en cuanto a lo que hoy podrían ser consideradas las bases fundamentales la convivencia, en una sociedad, manipulada o no, despojada o no de la posibilidad de acceder a contenidos veraces… que parece estar en una profunda crisis de identidad. No se ocultan las dudas que surgen acerca de si, en verdad, esta sociedad reclamaría y respaldaría intervenciones restrictivas del Derecho Penal en estos ámbitos; o de cómo enfocar estos problemas cuando puede que nuestra sociedad (o ciudadanos concretos) no quieran ser protegidos de los eventuales peligros que puedan derivarse del uso de las neurotecnologías, priorizando sus potenciales beneficios a corto o medio plazo; transitando hacia otro modelo y a otro contrato social sostenido por la inteligencia artificial...(Inglese/Lavazza, 2021) Una posición político-criminal restrictiva como la que aquí se mantiene… muy probablemente está, creo, condenada al fracaso.

6. BIBLIOGRAFÍA

Ansede, Manuel (2022), "Tener un sensor en la cabeza será de rigor en 10 años, igual que ahora todo el mundo tiene un teléfono inteligente" en: *El País*. https://elpais.com/ciencia/2022-01-05/tener-un-sensor-en-la-cabeza-sera-de-rigor-en-10-anos-igual-que-ahora-todo-el-mundo-tiene-un-telefono-inteligente.html

Drew, Liam (2019), "Agency and the algorithm", en: *Nature*, 571, 3, 19-21.

Gilbert, Frederic, Mathew Cook, Terence O'Brien, Terence, Judy Illes (2019), en: "Embodiment and Estrangement: Results from a First-in-Human 'Intelligent BCI' Trial", en: *Science and Engineering Ethics* 25, 1, 83-96. https://doi.org/10.1007/s11948-017-0001-5

Goering, Sara, Timothy Brown, Eran Klein (2021), "Neurotechnology ethics and relational agency", en: *Philosophy Compass*, 16, 4. https://doi.org/10.1111/phc3.12734

Goering, Sara, Eran Klein, Specker Sullivan, Rafael Yuste *et al.* (2021), "Recommendations for Responsible Development and Application of Neurotechnologies", en: *Neuroethics* 14, 365-386 https://doi.org/10.1007/s12152-021-09468-6

Fernando René González Moraga, Stéphani Klein Tuente, Sean Perrin *et al.*, (2022). "New Developments in Virtual Reality-Assisted Treatment of Aggression in Forensic Settings: The Case of VRAPT", en: *Frontiers in Virtual Reality*, 2, 675004. https://doi.org/10.3389/frvir.2021.675004

González-Tapia, María Isabel, Ingrid Obsuth, I (2015), "Bad genes" & criminal responsibility" en: *International Journal of Law and Psychiatry*, 39, 60-71. https://doi.org/10.1016/j.ijlp.2015.01.022

González-Tapia, Maria Isabel, (2013), "La concepción de la norma en un Estado Social y Democrático de Derecho. La dicotomía «valoración versus imperativo» en la tradicional discusión de la teoría de las normas penales", en: Hernández Romo Valencia, Pablo, Roberto Ochoa Romero, Luis Norberto Cacho Pérez (eds.) *Estudios Penales en homenaje al Profesor Javier Alba Muñoz,* Valencia, Tirant lo Blanch, 235-258.

Hildt, Elisabeth (2015), "What will this do to me and my brain? Ethical issues in brain-to-brain interfacing", en: *Frontiers in Systems Neuroscience*, 9. https://doi.org/10.3389/fnsys.2015.00017

Ienca, Marcello (2021), "On Neurorights", en: *Frontiers in Human Neuroscience*, 15, 701258. https://doi.org/10.3389/fnhum.2021.701258

Inglese, Silvia, Andrea Lavazza, (2021), "What Should We Do With People Who Cannot or Do Not Want to Be Protected From Neurotechnological Threats?", en: *Frontiers in Human Neuroscience*, 15, 703092. https://doi.org/10.3389/fnhum.2021.703092

Kaimara, Polyxeni, Andreas Oikonomou, Ioannis Deliyannis, (2022), "Could virtual reality applications pose real risks to children and adolescents? A systematic review of ethical issues and concerns", en: *Virtual Reality* 26, 697-735 https://doi.org/10.1007/s10055-021-00563-w

Klein, Eran, Sara Goering, S, Josh Gagne, Shea *et al.* (2016), "Brain-computer interface-based control of closed-loop brain stimulation: Attitudes and ethical considerations", en: *Brain-Computer Interfaces*, 3, 3, 140-148. https://doi.org/10.1080/2326263X.2016.1207497

Kögel, Johannes, Ralph J. Jox, Orsola Friedrich (2020), "What is it like to use a BCI? - Insights from an interview study with brain-computer interface users" en: *BMC Medical Ethics*, 21, 1, 2. https://doi.org/10.1186/s12910-019-0442-2

Maloney Divine (2021), *A Youthful Metaverse: Towards Designing Safe, Equitable, and Emotionally Fulfilling Social Virtual Reality Spaces for Younger Users*, All Dissertations Tigerprints.

Mazza, Massimiliano, Kornelius Kammler-Sücker, Tagrid Leménager *et al.* (2021), "Virtual reality: A powerful technology to provide novel insight into treatment mechanisms of addiction", en: *Translational Psychiatry*, 11, 1, 617. https://doi.org/10.1038/s41398-021-01739-3

Monaghesh, Elham, Taha Samad-Soltani, & Sara Farhang (2022), "Virtual reality-based interventions for patients with paranoia: A systematic review", en: *Psychiatry Research*, 307, 114338. https://doi.org/10.1016/j.psychres.2021.114338

Pugh, Jonathan (2020), "Clarifying the Normative Significance of 'Personality Changes' Following Deep Brain Stimulation", en: *Science and Engineering Ethics*, 26, 3, 1655-1680. https://doi.org/10.1007/s11948-020-00207-3

Rueda, John, Francisco Lara (2020), "Virtual Reality and Empathy Enhancement: Ethical Aspects", en: *Frontiers in Robotics and AI*, 7, 506984. https://doi.org/10.3389/frobt.2020.506984

Sapolsky, Robert M. (2018), "Doubled-Edged Swords in the Biology of Conflict", en: *Frontiers in Psychology*, 9, 2625. https://doi.org/10.3389/fpsyg.2018.02625

— (2021), "Neuroscience and the Law", en: *University of St. Thomas Journal of Law and Public Policy*, 138.

Savage, Neil (2019), "How AI and neuroscience drive each other forwards", en: *Nature*, 571, 7766, S15-S17. https://doi.org/10.1038/d41586-019-02212-4

Silva, Gabriel A (2018), "A New Frontier: The Convergence of Nanotechnology, Brain Machine Interfaces, and Artificial Intelligence", en: *Frontiers in Neuroscience*, 12, 843. https://doi.org/10.3389/fnins.2018.00843

Steinert, Steffen, Orsolya Friedrich (2020), "Wired Emotions: Ethical Issues of Affective Brain-Computer Interfaces", en: *Science and Engineering Ethics*, 26 (1), 351-367. https://doi.org/10.1007/s11948-019-00087-2

Yuste, Rafael, Goering, Sara, Blaise A. Arcas, Guogian Bi, José María Carmena *et al.* (2017), "Four ethical priorities for neurotechnologies and AI", en: *Nature*, 551 (7679), 159-163. https://doi.org/10.1038/551159a

Zeki, S., Goodenough, O. R., & Sapolsky, R. M. (2004) "The frontal cortex and the criminal justice system", en: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359 (1451), 1787-1796. https://doi.org/10.1098/rstb.2004.1547

Zuboff, Shoshana, (Albino Santos Mosquera, trad.) (2020). *La era del capitalismo de la vigilancia: La lucha por un futuro humano frente a las nuevas fronteras del poder,* Paidós, Barcelona.

CAPÍTULO XIV

REFLEXIONES EN TORNO A LA PERSONALIDAD ELECTRÓNICA DE LOS ROBOTS

ADOLFO J. SÁNCHEZ HIDALGO
Universidad de Córdoba
ji2sahia@uco.es

"Si se conceden derechos a los robots, tal vez la gente deje de comprarlos".
ISAAC. ASIMOV, *El hombre del Bicentenario*.

1. INTRODUCCIÓN

En 1972 Isaac Asimov publicó un pequeño relato de ciencia ficción titulado *El hombre del Bicentenario* en el que narraba los deseos de un androide por ser reconocido jurídicamente como persona y disfrutar de los mismos privilegios que sus compañeros humanos. Hoy, cincuenta años después, la comunidad académica celebra simposios y debates acerca de esta cuestión y otras muchas, que hace apenas algunas décadas parecían meramente fantásticas. Quizás pueden ser sólo quimeras de una nueva fe tecnológica, o quizá sean verdaderamente acontecimientos por venir, que cambiarán nuestra visión del mundo y de nosotros mismos. Aún no tenemos una plena conciencia de ello, pero, se nos pide un esfuerzo de anticipación a estos cambios y preventivamente disponer de los instrumentos necesarios para proporcionar una respuesta a los conflictos del futuro. Es probable que nuestras teorías resulten inútiles en esta empresa, pues como escribe Hegel (1993, 61) en el prefacio de sus *Fundamentos de Filosofía del Derecho*: "el búho de Minerva sólo levanta su vuelo al romper el crepúsculo". Porque, la sabiduría sigue el rastro de los acontecimientos. El presente está continuamente haciéndose y sólo en cuanto deviene pasado, podemos dominarlo. El futuro, en cambio, será o no, creemos alcanzarlo; pero son sólo ensoñaciones.

Hay muchísima literatura en los últimos 70 años dedicada a la IA y últimamente asistimos a un renovado interés como consecuencia de los avances en los sistemas de redes neuronales y el *Deep machine learning*, sin embargo, las actitudes frente a esta cuestión son esencialmente dos (Wilcocks, 2020, 287-288): un optimismo *hype*[1] y un pesimismo apocalíptico. En el primer caso se generan unas expectativas desmedidas sobre los beneficios de la robotización a consecuencia de una sobrevaloración de las capacidades de la IA. Por el

[1] *Hype*, se refiere a las expectativas generadas artificialmente alrededor de una persona o producto, cuya campaña promocional e imagen se ha construido a partir de la sobrevaloración de sus cualidades

contrario, los pesimistas dibujan un escenario distópico en el que la robotización acabará con los empleos y la vida humana tal como la conocemos (Fukuyama, Sandel, Richard y Daniel Susskind, etc).

Mi posicionamiento es escéptico, creo que debemos mantener los pies en la tierra, reflexionar sobre lo que sí sabemos acerca de la IA y los conflictos que está generando su progresiva implantación. Y, sobre todo, sobre todo, juzgar críticamente el coste y los beneficios de esta nueva revolución tecnológica para los derechos y libertades del hombre.

Nuestra situación es la siguiente: perdida la batalla en el frente tecnológico, donde USA y China capitalizan prácticamente toda la innovación en este ámbito; la UE se ha decidido por dominar el frente ético y jurídico, por lo que ha abordado desde una perspectiva total los problemas que los modernos sistemas de IA y robots autónomos puedan generar en la sociedad y cómo pueden afectar a los derechos de los ciudadanos. Entre la múltiple documentación normativa de carácter comunitario, deben destacarse: la Directiva 1999/34/CE del Parlamento europeo y del Consejo de 10 de mayo de 1999, la Directiva (UE) 2019/771 de 20 de mayo de 2019, , el Dictamen 19/09/2018 del Comité Económico y Social Europeo sobre «Inteligencia artificial: anticipar su impacto en el trabajo para garantizar una transición justa», y, en especial, la reciente Propuesta 21/04/2021 de Reglamento del Parlamento europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión.

Sin duda, una empresa encomiable; pero con el riesgo de que esta imparable revolución tecnológica nos pase por encima. Porque mientras Europa está reunida en comisiones y demás eventos de todo tipo discutiendo sobre el futuro de las máquinas; en Asia habrá alguien en un laboratorio o fábrica haciéndolo realidad.

La cuestión que nos ocupa es la personalidad electrónica de los robots y surge dentro de la Unión Europea como consecuencia de una sobrestimación de la autonomía de los robots "inteligentes", reflejada en las Recomendaciones de la Comisión de Asuntos Jurídicos a la Comisión sobre normas civiles en materia de Robótica, publicada en enero 2017, concretamente en su considerando 59. f): "Crear a largo plazo una personalidad jurídica específica para los robots, de forma que como mínimo los robots autónomos más complejos puedan ser considerados personas electrónicas responsables de reparar los daños que puedan causar, y posiblemente aplicar la personalidad electrónica a aquellos supuestos en los que los robots tomen decisiones autónomas inteligentes o interactúen con terceros de forma independiente".

Esta iniciativa fue rápidamente contestada en el plano internacional por el Informe Unesco 14 septiembre de 2017 a propósito de una ética sobre los robots en el que se desaconsejaba este tipo de iniciativas (punto 201 del Informe)[2]; aunque, también, dentro de los propios especialistas europeos como Nathalie Nevejans (CNRS) que lideraron un movimiento en contra de esta posibilidad (*Robotics open letter*, http://www.robotics-openletter.eu/), argumentando que se sobrestimaban las capacidades de los robots y reflejaban un conocimiento superficial de los procesos de aprendizaje y toma de decisiones por parte de los robots, incluso los más avanzados.

Como consecuencia de estas reacciones y posiblemente también por otras causas que desconocemos, lo cierto es que en los sucesivos instrumentos normativos de la UE no hay una sola mención a la personalidad jurídica de los robots, ni en la Directiva (UE) 2019/771, ni en las orientaciones éticas para una IA confiable (2019) dadas por el Comité de expertos en IA; ni en las conclusiones de la Presidencia sobre la incidencia de la IA en los derechos fundamentales de la Unión (21 octubre 2020); como tampoco en los resultados de los *hubs*, *cluster* o lobbies en esta materia (proyecto ADRA 2021). Es más, hace apenas unos meses fue aprobado por la UNESCO un documento con Recomendaciones sobre ética de la IA y en su recomendación 68 se rechaza el reconocimiento de personalidad jurídica a los robots[3]. En suma, la política de la UE es priorizar la protección de los derechos y garantías que pueden verse amenazados como consecuencia de los avances en IA.

Desde una perspectiva dogmática el tema de la personalidad jurídica de los robots ha sido objeto de todo tipo de propuestas, además, no debe descuidarse el hecho de que el concepto de persona ha sido un tema central del pensamiento jurídico a lo largo de la historia. Entre las propuestas más originales destacaré la posición de Ercilla (2018a, 6.5), quien encuentra en la

[2] However, it is highly counterintuitive to call them 'persons' as long as they do not possess some additional qualities typically associated with human persons, such as freedom of will, intentionality, self-consciousness, moral agency or a sense of personal identity.

[3] 68. Los Estados Miembros deberían elaborar, examinar y adaptar, según proceda, marcos reguladores para alcanzar la rendición de cuentas y la responsabilidad por el contenido y los resultados de los sistemas de IA en las diferentes etapas de su ciclo de vida. Cuando sea necesario, los Estados Miembros deberían introducir marcos de responsabilidad o aclarar la interpretación de los marcos existentes para garantizar la atribución de la responsabilidad por los resultados y el funcionamiento de los sistemas de IA. Además, al elaborar los marcos reguladores, los Estados Miembros deberían tener en cuenta, en particular, que la responsabilidad y la rendición de cuentas deben recaer siempre en última instancia en personas físicas o jurídicas y que no se debe otorgar personalidad jurídica a los propios sistemas de IA. Para lograrlo, esos marcos reguladores deberían ajustarse al principio de la supervisión humana y establecer un enfoque global centrado en los actores de la IA y los procesos tecnológicos que intervienen en las diferentes etapas del ciclo de vida de los sistemas de IA.

figura del esclavo romano las claves para una posible regulación de la personalidad jurídica de los robots, o como él prefiere llamarla, personalidad ciber-física. En otras direcciones hay autores que intentan extender analógicamente la idea de personalidad a los robots a razón de su similitud con el hombre (Zimmerman, 2015, 43); o bien, mediante el empleo de la misma ficción que se utiliza para atribuir personalidad a las corporaciones y sociedades (Broman & Finckerberg-Broman, 2018, p. 73); o, incluso, la de reconocer subjetividad jurídica como ocurre con ciertos patrimonios, como un estatuto intermedio entre personalidad y propiedad (Nagenborg, 350). Por supuesto, también hay autores que niegan de todo punto esta personalidad jurídica, ya sea por razones dogmáticas o pragmáticas (Herrera, 2022, 48-49).

Ante esta diversidad de posiciones, creo que lo más correcto teóricamente es tomar distancia y abordar el problema de la idea de persona en su complejidad, es decir, asumir la pluralidad de perspectivas desde las que se puede afrontar su conocimiento y entre las cuales la visión jurídica es sólo una de ellas.

2. LA PERSONALIDAD DE LOS ROBOTS DESDE UNA PERSPECTIVA RELIGIOSA

En las principales religiones del mundo, las denominadas religiones del libro (judaísmo, cristianismo e islam), no cabe duda de que el ser humano es un producto de la creación divina y situado en una posición privilegiada respecto al resto de seres de la creación. En la teología cristina el hombre se asemeja a Dios (*Imago Dei*), pero no es Dios (De Aquino, 2001, 321-326). En este sentido, Echevarría (2019, 30) explica que para Tomás: "en el hombre se llama la imagen de Dios por su naturaleza racional, ya que por su intelecto imita la naturaleza divina de una manera que refleja en cierto modo su especie. Pero esa representación de la especie no ha de entenderse como si Dios y el hombre se colocaran en la misma especie ni género. Es una representación analógica, no unívoca. El ser humano participa de la divinidad; pero no le es dado conocer todos los misterios del universo, que competen sólo a Dios. Desde la perspectiva cristiana no habría ninguna duda en negarle dignidad a la máquina, porque no participa de la condición *imago Dei*, ni siquiera de la condición *imago homine*, porque no hay ningún parentesco entre la realidad orgánica y fisiológica que es el hombre y la realidad mecánica del autómata (Wiener, 1966, 12).

En el caso de la religión islámica la persona humana es sólo el creyente monoteísta que debe convertir al mundo entero, por ello el concepto de persona se identifica con el testimonio de fe y sólo en la medida en que el creyente da testimonio de la Verdad -en el culto, ritos y virtud- se asemeja o, en su rostro, se ve la imagen de Dios (Massignon, 1952, 451-453). Como ha explicado Lahabi (1967, 47-63), el concepto de persona tiene una doble dimensión física y

axiológica, pero es la profesión de fe la que dignifica el rostro o dimensión física del hombre. En consecuencia, parece improbable la posibilidad de reconocer en el robot a una persona, en tanto carece de esta espiritualidad.

Sin embargo, no debe menospreciarse el hecho de que lo que denominamos Inteligencia Artificial es, en suma, un código de programación o, de otro modo, un mensaje que no sólo moldea la máquina; sino que también puede ser reproducido mecánicamente en múltiples artefactos (Wiener, 1966, 36). En este sentido, si la máquina fuera un medio de transmisión del testimonio y la virtud divina, quedaría por ello dignificada y se asemejaría a la imagen de Dios. Porque no es la realidad física lo que define la persona humana, sino su realidad espiritual (Bellver, 2013, 65). Ahora bien, es más que dudoso que eso ocurra, por muy avanzados que sean los sistemas de aprendizaje o entrenamiento neuronal de los sistemas de IA, ya que el paradigma de IA es una inteligencia técnica-operacional. Única inteligencia reducible al lenguaje formalizado del algoritmo o de la programación.

Aunque no faltan noticias del empeño del hombre por crear un ser humano por medios diferentes a la generación natural, lo cierto es que sólo a través de nuestra naturaleza biológica somos capaces de crear nuevos seres humanos. Quizás el judaísmo sea la religión en la que más se ha abordado este asunto (Ben-Naftali & Triger, 2013, 25-27). Siguiendo el estudio de Trachtenberg (2004, 84-103) se observa que el Talmud reconoce una segunda forma de creación, que requiere la aplicación de la Leyes de la Creación, probablemente una colección de tradiciones místicas relacionadas con la creación original del universo. Este tipo de magia era la única permitida al inicio, por medio de ella si el justo lo desea puede crear un universo. Se nos dice que Raba creó un hombre y lo envió al rabino Zeira, que conversó con él pero como no respondía, exclamó: Has sido creado con magia, vuelve al polvo. Rabbis Hanina y Oshaya solían reunirse los viernes y discutir sobre el Libro de la Creación y en unas de sus reuniones supuestamente crearon un cordero que se comieron.

Rashi describe el método que la tradición había preservado a través de sus comentaristas: ellos suelen combinar las letras del Nombre por el que el universo fue creado; esto no se considera magia prohibida, pues los trabajos de Dios conforman la existencia a través de su Nombre. Judíos y cristianos medievales estaban deseosos de crear vida humana y creían en la habilidad del hombre para hacerlo. Guillermo de Auvergne escribió: "los hombres han intentado producir y pensado en el éxito de crear vida humana de otros modos diferentes al proceso generativo usual", pero los métodos empleados por los no judíos, son menos sutiles que los propuestos en el Talmud. Por ejemplo, en el siglo XIV un escrito cita al árabe Rasis (siglo X), quien creó un ser humano poniendo una sustancia innominada en un bote de estiércol de caballo durante

tres días. En el siglo XIII el alemán Hasidim (pietista y místico) estaba especialmente intrigado en este asunto, de él proviene el uso de la palabra *golem* (literalmente, amorfo o carente de vida) para designar al homúnculo creado por la mágica invocación de nombres. El primer individuo que parece haber creado uno, fue el rabino Samuel padre de Judah el piadoso, pero no podía hablar. Joseph Delmedigo nos informa, en 1625, que existen muchas leyendas de este tipo en Alemania y quizás debamos creerle. Entre las más conocidas está una relacionada con Elijah de Chelm (mediados del siglo XVI) y que fue contada durante el siglo XVII. A él se le atribuye crear un *golem* de arcilla a través del Libro de Yezirah, inscribiendo el nombre de Dios en su frente, cuando este ser cobró vida no tenía la capacidad de hablar. Al adquirir la criatura un tamaño gigantesco y una enorme fuerza, eliminó el nombre de su frente y se convirtió en polvo.

Con más detalle, un místico alemán, Eleazar de Worms describe la fórmula de su creación: la imagen debía de estar formada por arena de una montaña no pisada antes por el hombre y el encantamiento comprendía el alfabeto de las 221 puertas, que debía ser recitado sobre cada órgano. Un último detalle consistía en inscribir el nombre de Dios o la palabra hebrea "Emet" (verdad). La destrucción de la criatura devenía borrando la letra inicial del nombre, dejando la palabra hebrea "Met", que significa muerte. Como recuerda Jacob Shalom, judío llegado a Barcelona en 1325 proveniente de Alemania, la ley de la destrucción es el reverso de la ley de la creación.

Aunque por lo general, los judíos medievales eran bastante escépticos sobre la posibilidad de infundir vida a materias inertes, reconocían que la manipulación de los nombres respondía a un propósito más alto que ellos mismos. En 1625 sarcásticamente Zalmanteví de Aufenhausen reconoce: "nuestros tontos no son de arcilla, sino que provienen del vientre de sus madres". La actitud general de los judíos sobre este tema es reconocer que se puede hacer vida artificial, pero se han perdido los medios y es mejor así. La creación de estos homúnculos es vista como un acto de hechicería en el que las consecuencias son por lo general igualmente de imprevisibles que perniciosas.

Desde el punto de vista religioso, la cuestión parece clara, ya sea a través de la clonación o por la robótica, la creación de un ser artificial sería usurpar la posición de Dios y sembraría la semilla de nuestra destrucción. Wiener (1966, 52), incluso, nos señala el peligro de una nueva forma del pecado de simonía[4], esta es, cuando se usa la magia de la automatización para el enriquecimiento

[4] Recordar tan sólo al lector que el pecado de simonía consistía en usar los sacramentos y el poder espiritual con fines lucrativos o espurios.

personal y la lucha por el poder, en lugar de contribuir a mejorar la condición humana.

Leyendo entre las líneas de nuestros textos sagrados, a pesar del deber de sumisión y devoción a Dios, existe una pulsión irrefrenable del hombre a alzarse contra su creador. ¿Por qué habría de ser diferente en el caso de las máquinas si llegaran a ser realmente humanas? De llegar el profetizado momento de la "singularidad"[5], quién puede asegurar que este robot no nos verá como nosotros vemos a los simios y demás parientes evolutivos. Recuérdense las palabras del Zaratustra de Nieztsche (1965, 244): "¿Qué es el mono para el hombre? Un motivo de risa o una dolorosa vergüenza".

3. LA CONDICIÓN DE PERSONA DE LOS ROBOTS DESDE UNA PLANTEAMIENTO FILOSÓFICO-ANTROPOLÓGICO

Desde una perspectiva filosófica la noción de persona aparece directamente vinculada a la condición racional del ser humano, es decir, a su racionalidad y voluntad, que nos elevan respecto del resto de seres vivos. San Agustín dividía al hombre en hombre exterior y hombre interior, el exterior en nada se diferencia del resto de los animales; sin embargo, el hombre interior evoca la imagen (persona) de Dios y su alma es análoga a la Trinidad: la memoria es reflejo del Padre, la inteligencia del Hijo y la voluntad es la síntesis de ambas, el Espíritu santo (Morote Sarrión, 2009, 29-32). En suma, la imagen o persona moral, que eleva al hombre, es su voluntad.

Santo Tomás (2001, 320) siguiendo a Boecio lo define más profundamente como "sustancia individual de naturaleza racional". El de Aquino destaca, entonces, no sólo la nota de racionalidad de la criatura humana; sino, también, su dignidad como individuo, es decir la unicidad de cada ser humano. Es la condición racional, se le llame naturaleza en Tomás o forma en Aristóteles, la que permite al ser humano tomar posición de su individualidad y obrar libremente, sin estar movido por otra sustancia.

Inmanuel Kant (2008, 30) considera sinónimos hombre y persona[6], más aún, al definir al hombre como realidad nouménica o trascendental, destaca la

[5] Por singularidad en IA se entiende el hipotético acontecimiento en el futuro, por el cual se producirá una explosión de inteligencia en la máquina que dejará muy atrás a la humana. Esto se considera reforzado por el hecho de la explosión de velocidad en el rendimiento de los sistemas de IA, de modo que esta explosión inteligente será más pronto de lo que podamos imaginar (Chalmers, 2010, 7-8).

[6] "Persona es el sujeto, cuyas acciones son imputables. La personalidad moral, por tanto, no es sino la libertad de un ser racional sometido a leyes morales (sin embargo, la psicología es únicamente la facultad de hacerse consciente de la identidad de sí mismo en los distintos estados de la propia existencia), de donde se desprende que una persona no está sometida a otras leyes más que las que se da a sí misma (bien sola o, al menos, junto con otras)".

idea de autonomía moral; esto es, su aptitud para determinar libremente el contenido de su voluntad. Una tarea en la que la racionalidad y la conciencia de los demás cumplen una función orientadora (Kant, 2008, 30-32).

De acuerdo con esta perspectiva filosófica, deberíamos preguntarnos si los robots inteligentes que conocemos comparten estos atributos. La respuesta es obvia, no; por lo que en ningún caso pueden ser considerados como personas en este sentido. Entiéndase, sin perjuicio de que puedan ser utilizados para la toma de decisiones morales, usando su capacidad de procesamiento de datos para medir las consecuencias de nuestras decisiones y fortalecer así nuestra moralidad (Savulescu, & Maslen, 2015, 83). Claro que deberíamos ser cautelosos y no entregar nuestras decisiones a estos sistemas de IA sin conocer debidamente las complejidades de sus procesos computacionales. Como afirmaba Wiener (1966, 69) hay muy poca esperanza de que estos esclavos electrónicos nos ofrezcan un mundo en el que el hombre pueda dedicarse a la vida contemplativa; al contrario, lo más probable es que el mundo del futuro amenace y ponga de manifiesto las limitaciones de nuestra inteligencia.

Es ciertamente difícil concebir a los robots como agentes morales, tanto en un sentido positivo como negativo. Difícil es concebir la posibilidad de que un robot experimente el sentido del deber y el sentimiento de culpa, lo que constituye el fundamento de la imperatividad ética-moral. El sentido del deber y la culpa, en tanto humanos, requieren en el sujeto la capacidad de autodeterminación o libertad moral del hombre, para la elección de los fines y la ordenación de su voluntad. La máquina no puede actuar más que como su programación le ordena, actúa movida por su programación y no por sí misma, como afirma Moro (2015, 17) la subjetividad no se puede medir ni programar a través de la inteligencia algorítmica. En lenguaje jurídico sería muy inverosímil hablar de conducta dolosa en la máquina (Capellini, 2019, 512).

Así mismo, es realmente complicado concebir que el robot pueda ser sujeto de castigos morales, porque si no es capaz de sentirse culpable tampoco experimentará el arrepentimiento y la constricción. Tampoco tiene mucho sentido predicar la posibilidad del castigo físico en la medida en que carece de sistema nervioso y no puede sentir dolor, ni miedo a la muerte, como tampoco miedo a la reclusión (Asaro, 2012, 182).

Podemos aventurarnos más y preguntarnos qué ocurriría si los robots alcanzasen esta autonomía moral, esa unicidad que define al ser humano. Sinceramente considero improbable el advenimiento del fenómeno de la singularidad. En suma, porque la denominada Inteligencia Artificial no deja de ser una racionalidad operativa fundada en una comprensión formalizada (algorítmica) de nuestro entorno; y no parece posible que alcancen lo que los filósofos clásicos denominaban racionalidad práctica, o lo que los filósofos de la

modernidad, de otro modo, denominan el lenguaje ético de los fines (Rodríguez Puerto, 2021, 89).

Como afirma Polo (1997, 228): "el hombre no hace nada sin que al hacerlo no se produzca alguna modificación de su propia realidad espiritual", lo que significa que con cada operación se produce una retroalimentación o hábito intelectual, en virtud del cual se reformula nuestra visión del bien a perseguir y su naturaleza. De ese modo el conocimiento ético (a excepción del hábito de los primeros principios) no nos viene impuesto por la razón como un a priori; sino que, por el contrario, la razón lo descubre en cada decisión y en el ejercicio de la libertad intrínseca al ser humano (Polo, 1997, 228-237). Ciertamente, sería posible introducir una congenie de principios éticos, máximas morales y bienes éticos que presidieran una IA construida a imagen de la humana (Yue-Hsuan, Chien-Hsun, Chuen-Tsai, 2009, 275); pero es improbable que, en el curso de sus acciones, llegara a redefinir o reprogramar estos parámetros. Y, aún antes, este catálogo de datos éticos sería una formalización banal y grotesca de la realidad circunstancial en la que se incardina el hombre como criatura moral. ¿Por último, qué ocurriría con las emociones que están en la base de nuestra moralidad, en qué medida podrían ser introducidas en la mente de la máquina? Debe señalarse con objetividad que la persona es algo más que una visión funcional u operativa, implica características personales y biológicas como la inteligencia, el conocimiento, los sentimientos y todo ello interconectado (McFee, 2019, 191). La moderna teoría cognitiva de la neurociencia ha demostrado que las emociones son componentes de los procesos neuronales que inciden en el acto mismo de conocer y también en la toma de decisiones, de modo que nuestra racionalidad está íntimamente vinculada a nuestras emociones y sólo desde esta premisa podemos comprender la particularidad del conocimiento y la acción humana (Fuselli, 2018, 107).

Ahora bien, si fuese el caso de que, aún muy limitadamente y sin alcanzar del todo la problematicidad de lo humano, adquiriesen cierta racionalidad ética semejante a la humana, habrían de considerarse ciertamente personas. ¿No lo haría el lector, si ahora apareciera un extraterrestre semejante a nosotros? En nuestro recuerdo está la famosa escena del *Planeta de los Simios* en la que un tribunal de tres simios juzga si Charlton Heston debe ser reconocido como sujeto de derechos. Todos nos hemos situado a su favor y lo hacemos no porque sea humano, sino porque es lo justo. Si se reconociese en el robot esa compleja racionalidad del humano, por extraño que nos resultase deberíamos estar abiertos a considerarlo algo más que una cosa o una herramienta (Zimmerman, 2017, 21).

4. LA CONDICIÓN DE PERSONA ARTIFICIAL DESDE UNA VISIÓN NEUROLÓGICA

Esta perspectiva es de especial interés, puesto que la idea de Inteligencia Artificial encuentra su origen en la posibilidad de simular la inteligencia humana. La idea de crear una máquina con la capacidad de reproducir el cerebro humano comenzó a gestarse a principios de 1940, cuando McCuloch y Pitts describieron el primer modelo computacional de una neurona artificial y, con ello, abrieron el horizonte científico de las llamadas redes neuronales artificiales. La hipótesis de partida es muy básica, si se consigue reproducir artificialmente el cerebro humano, en toda su complejidad funcional y en sus múltiples interconexiones neuronales, se habrá conseguido replicar la inteligencia humana (Daley, 2018, 36).

Dos dificultades presenta esta idea. La primera es de carácter técnico y previsiblemente podrá ser superada en el futuro: ahora mismo los avances técnicos no permiten reproducir los millares y millares de conexiones neuronales de nuestro cerebro. Además, a esta dificultad debe añadirse el hecho de que aún los neurólogos no han alcanzado un completo conocimiento del cerebro humano y quedan décadas de investigación para alcanzarlo. La consecuencia es obvia, no se puede reproducir lo que no se conoce (Searle, 2000, 170-172). En este punto, los más optimistas consideran que quizás lo que el hombre no alcanza a conocer, puede ser conocido con más prontitud por los modernos sistemas de IA al servicio de las investigaciones científicas. Una situación bastante irónica, porque las máquinas adquirirían un conocimiento total del cerebro humano antes de que el hombre se conozca a sí mismo.

La segunda dificultad es atinente al modelo de racionalidad científica con el que se aborde el problema del cerebro humano. Los teóricos de la computación y la IA conciben el cerebro humano en un sentido marcadamente empirista (a imagen del modelo trazado por David Hume), de modo que el conocimiento humano no es sino la suma de las experiencias individuales, en una especie de representación fantasmal en el teatro de nuestro cerebro y que trae como resultado la formación de lo que denominamos "yo" o "conciencia". Según esta concepción, la individualidad o el "yo" no es dado a priori, se forja como una especia de contaminación ligada a nuestras acciones y percepciones sensitivas (Popper & Eccles, 1985, 115-117). De acuerdo con esta perspectiva, los sistemas de IA o los robots autónomos del futuro (la llamada superinteligencia artificial) podrían llegar a desarrollar un "yo" o "conciencia", como resultado de la suma progresiva de sus experiencias cognitivas (Hubbard, 2011, 446).

Esta posición es muy discutible, primero porque todavía no tenemos la más mínima idea de qué sea la conciencia humana, ni por supuesto, cuál es el origen del yo. La posición empirista es tan sólo una hipótesis y no parece la más

acertada entre el conjunto de posiciones teóricas. No es ningún secreto para la neurociencia que el cerebro humano es un gran laboratorio químico y que cada conectoma humano es radicalmente único, esto es, las reacciones químicas que despierta el movimiento neuronal en nuestro cerebro es un fenómeno estrictamente individual e irrepetible. Apenas llegamos a conocer las principales hormonas que operan en nuestro cerebro, los hemisferios en los que actúan y sus implicaciones potenciales en el carácter. Una explicación muy seguida en este ámbito científico es la ofrecida por MacLean, que divide el cerebro humano en tres córtex y el modo en que se interrelacionan definiría nuestra subjetividad o conciencia, nuestra inteligencia y nuestra percepción espacio-temporal (el córtex protoreptiliano[7] -primitivo-, el córtex mamífero[8] -paleopalino- y el córtex sapiens[9] -neopalino-). Ahora bien, el hecho es que la razón de la formación de estos tres niveles del cerebro humano es esencialmente evolutiva y el modo en que interactúan entre sí es biológico, esto es, mediante reacciones químicas (MacLean, 1990). Es precisamente en esta interacción o conflicto entre conectoma (red neuronal) individual, la química cerebral y la estructura trina del cerebro, donde se encontraría la explicación de la singularidad del cerebro humano: la moralidad y la espiritualidad del ser humano (Cory Jr, 2000, 406). Más aún, debe reconocerse que el yo, la conciencia, la moralidad y la espiritualidad son productos tanto de la biología individual como de la inserción social del individuo y sus experiencias, por lo que la cuestión resulta aún más compleja. En definitiva, esta multiplicidad de elementos biológicos, químicos, sociales y sus interacciones quedan finalmente reducidos a la afirmación de la radical unicidad de todo ser humano.

A diferencia de lo anterior, el modo en que se ha proyectado la IA es estrictamente mecanicista. Los teóricos de la IA distinguen tres niveles cognitivos: la acción inteligente, la inteligencia autónoma y la inteligencia humana simulada. La acción inteligente sería el equivalente a nuestro sistema nervioso: coordina percepciones y comportamiento; la inteligencia autónoma implicaría la capacidad para resolver problemas, mediante la identificación de patrones, planificación de acciones y aprendizaje autónomo; la inteligencia humana simulada implicaría la capacidad para el razonamiento abstracto y la formación de una conciencia o mente propia (Yue-Hsuan, Chien-Hsun, Chuen-Tsai, 2009, 274). El desarrollo conjunto de estas tres dimensiones

[7] En el córtex protoreptiliano se situaría el instinto de autoconservación y supervivencia.

[8] En el córtex mamífero se encontraría el deber de cuidado y solidaridad con los semejantes, sería la fuente del sentimiento de empatía y condolencia propio del hombre.

[9] El córtex sapiens es exclusivo del hombre y surge evolutivamente como consecuencia de la interacción entre el córtex primitivo y paleopalino, dando lugar a la autoconsciencia, la conciencia moral y el sentido de trascendencia.

conduciría a la idea de IA fuerte, que ha sido definida por García (2020, 85) como: "aquel hipotético sistema capaz de emular el funcionamiento de la mente humana, incluyendo no solo la capacidad de resolución de multitud de tareas sino, también, los sentimientos, la creatividad y la auto-conciencia".

Se ha avanzado muchísimo en las últimas dos décadas en la acción inteligente y en la inteligencia autónoma, aunque la posibilidad de una inteligencia humana simulada es todavía una quimera. Los más avanzados sistemas de IA (sistemas de redes neuronales y *Deep machine learning*) son capaces de percibir el entorno y proyectar acciones; pero no comprenden el entorno, no son capaces de juzgar el acontecer, esto es, no poseen una cosmovisión. Es más, los proyectos de mejora del rendimiento o entrenamiento de los sistemas *Deep machine learning* están incidiendo en estos extremos, tratando de mejorar la capacidad de razonar y aprender mediante: a) un razonamiento que conjugue los valores geométricos y semánticos, permitiendo un reconocimiento de contextos al sistema; y b) un aprendizaje activo que permita la integración de incertidumbres, probabilidades en el flujo de datos del sistema (Sünderhauf *et al.*, 2018, 406-409). Cierto que en no pocas situaciones el éxito de estos procesos de aprendizaje y de las diferentes actividades para las que se ordenen, dependerá del empleo de criterios éticos, pautas de valor y juicios críticos que están muy lejos de poder ser traducidos formalmente en un lenguaje algorítmico (Wiener, 1966, 77).

Debe recordarse que el cerebro humano conoce muy poco sin la intervención del intelecto, sólo evidencias sensitivas. El resto de conocimientos implica una experiencia cognitiva propia y filtrada por nuestra potencia racional. Todos portamos una propia cosmovisión o horizonte de sentido -en continuo proceso de formación- desde el que comprendemos el acontecer y dirigimos nuestra conducta. Pero, la clave es que esta cosmovisión u horizonte semántico sería la prueba más tangible de la sinestesia y la plasticidad cerebral (Arenas, 2016, 1015) y dudo que sea posible reproducir o emular este contexto hermenéutico a través de unos patrones semánticos formales.

En suma, desde una perspectiva neurocientífica queda mucho camino por recorrer para que el "cerebro" de los robots emule al hombre, además esta meta tiene otras muchas vías por explorar. Se cumpliría aquí la máxima del Salmo 139: versículo 17: "Dios mío. ¡Qué difícil me resulta entender tus pensamientos! ¡Pero más difícil todavía me resultaría contarlos!"

5. LA PERSONALIDAD JURÍDICA DE LOS ROBOTS

Como se ha indicado en la introducción del estudio, el reconocimiento de personalidad jurídica a los robots es un asunto que se ha abandonado en los textos normativos del continente, a pesar de la importante discusión académica

que ha reportado en el último lustro. Hoy la preocupación comunitaria se centra en el establecimiento de unos principios éticos desde los que proyectar una IA confiable, la delimitación de lo que deba entenderse por Inteligencia Artificial y la incidencia de la IA en los derechos fundamentales de la Unión (Lacruz, 2020, 161-164). Son reveladoras de la nueva voluntad comunitaria, las palabras de la Presidencia del Consejo de Europa empleadas al inicio de las conclusiones sobre la Carta (2020) de los Derechos Fundamentales en el contexto de la Inteligencia Artificial y el Cambio digital: "Nuestro compromiso es que la IA se diseñe, desarrolle, despliegue, utilice y evalúe de manera responsable y centrada en el ser humano. Debemos aprovechar el potencial de esta tecnología clave para promover la recuperación económica en todos los sectores, con un espíritu de solidaridad europea, defender y promover los derechos fundamentales, la democracia y el Estado de Derecho y mantener unas normas jurídicas y éticas exigentes".

Si bien, esta situación no excluye la necesidad de abordar la cuestión de la personalidad jurídica de los robots; más aún, permite al académico tomar distancia y examinarlo con mayor rigor al no acusar las prisas del pragmatismo. En este sentido, de acuerdo con una visión jurídica material del concepto de persona no cabe duda de su antropomorfismo, esto es, el Derecho ha reconocido tradicionalmente como personas a entes o entidades semejantes al ser humano o, respecto de las que se predican las mismas cualidades que al ser humano. Quizás el caso más polémico sea el de persona jurídica, dada la tradición dogmática de considerarlas persona *ficta* o ficciones jurídicas en razón de su utilidad para el tráfico jurídico (Núñez, 2018, 20-21). No obstante, como recordaba Federico de Castro y Bravo (1949, 1414-1415) esta ficción sólo se sostiene por el hecho de que haya una pluralidad de seres humanos que la conforman y cuanto más se descuida el elemento subjetivo de la persona jurídica más se desdibuja o deforma su concepto.

Teniendo en cuenta que falta por la fuerza de las cosas la condición de subjetividad o el elemento humano, hay quien con espíritu dogmático apela al abismo infranqueable entre sujetos y objetos. Los robots serían en todo caso "cosas" y no deben ser tratados como sujetos (personas). Destaca en esta línea las tesis de Borelle (2018), Frison-Roche (2017) y la del propio Loiseau (2018) que califica la personalidad jurídica de los robots como "monstruosidad jurídica". Más eclécticamente, Larregue (2019) apela a la categoría de "bienes semovientes" del derecho romano para proporcionar nuevas direcciones en torno a la personalidad de los robots. Esta sería la dirección seguida por Ercilla (2018b, 22-25) en España, quien rescata la idea de los *status* en el derecho romano para defender la posibilidad de diversos niveles de subjetividad en el Derecho.

Por atractiva que pueda resultar la idea, lo cierto es que nuestros ordenamientos jurídicos continúan considerando a los robots como cosas, objetos o productos carentes absolutamente de personalidad (Asís, 2018, 58).

Claro que, desde una perspectiva estrictamente formal, no hay duda de que el concepto de persona para el Derecho es una creación del ordenamiento jurídico y como afirmaba Kelsen (2011, 75) no significa otra cosa que ser el centro de imputación de acciones, derechos y obligaciones. Es decir, el Derecho puede atribuir la condición de persona a cualquier entidad o patrimonio, lo que hace es construir una unidad dentro del ordenamiento jurídico, en la que se contienen una serie de derechos y obligaciones. Desde esta perspectiva formalista no habría ningún inconveniente en atribuir personalidad jurídica a los robots autónomos; pero, aunque sea sólo por prudencia deberíamos preguntarnos ¿para qué? ¿Cuál o cuáles serían las razones de su utilidad? No debe descuidarse la función pragmática o instrumental del concepto de persona, esto es, debe juzgarse en qué medida facilita la vida del hombre y le permite la consecución de sus fines, porque el Derecho es una herramienta al servicio del hombre (Medina, 2013a, 627). La cuestión formal, quedaría -como poco- subordinada a la cuestión pragmática. Llegados a este punto emerge la cuestión fundamental ligada a la personalidad jurídica de los robots: el problema de la responsabilidad jurídica de los mismos.

5.1. Personalidad jurídica y responsabilidad de los robots

El asunto de la responsabilidad es, en efecto, la cuestión principal detrás del debate de la personalidad jurídica de los robots, y, en suma, se busca ofrecer una respuesta al problema de los daños que éstos pudieran producir.

Hoy por hoy este problema está resuelto por medio del sistema de responsabilidad por daños derivados de productos defectuosos, que en el derecho comunitario tiene su origen en la Directiva 85/374/CEE del Consejo de 25 julio 1985 relativa a la aproximación de las disposiciones legales, reglamentarias y administrativas de los Estados miembros en materia de responsabilidad por los daños causados por productos defectuosos. Esta norma, se ha visto completada recientemente con una nueva Directiva (UE) 2019/771 de 20 de mayo de 2019 relativa a determinados aspectos de los contratos de compraventa de bienes, por la que se modifican el Reglamento (CE) n. 2017/2394 y la Directiva 2009/22/CE y se deroga la Directiva 1999/44/CE; que si bien no aborda por completo el tema, si introduce ya nuevas medidas aplicables a los sistemas de IA y servicios digitales. Estas nuevas medidas se caracterizan por permitir la formación de un litisconsorcio pasivo necesario entre usuario del producto, vendedor del producto, desarrollador y fabricante.

El problema de la responsabilidad por daños derivados de productos defectuosos, al menos en España, como ha puesto de manifiesto Asís (2018, 55), es que la legislación contiene una vía para la elusión de responsabilidad no indiferente, el artículo 140 del Real Decreto 1/2007, principalmente el de su letra e) "que el estado científico técnico del conocimiento existentes en el momento de la puesta en circulación del producto no permitía apreciar la existencia del defecto". Pero, en general, todas las excepciones del artículo 140[10] serían problemáticas en el caso de ser aplicadas a los robots o sistemas de IA.

La mayoría de los autores, en cambio, aboga por el establecimiento de un régimen de responsabilidad objetiva para estos supuestos, que se vea complementado con el establecimiento de una modalidad de seguro obligatorio para los usuarios de robots autónomos y, además, un Fondo de compensación de seguros. De este modo, se garantizaría la agilidad y la efectiva reparación de los daños producidos por las acciones realizadas por robots autónomos (Asís, 2018, 60-62).

Se discute, por otro lado, si el tipo de responsabilidad objetiva debe emular el modelo de la responsabilidad vicarial de los padres sobre las acciones de sus hijos, o de la empresa respecto de sus empleados; o, de otro modo, debe asemejarse al de los titulares de vehículos a motor (Lacruz, 2020, 206-220). A mi parecer, sería más apropiado el modelo de responsabilidad objetiva de los vehículos a motor, porque ontológicamente supone concebir como objetos o cosas a los robots autónomos y justificaría más coherentemente el sistema de seguro obligatorio y el Fondo de compensación de seguros. La dificultad reside en juzgar si el obligado a contratar el seguro de responsabilidad civil deber ser el usuario del robot autónomo, o, por el contrario, el fabricante o importador, suministrador del producto. Esta segunda opción sería más coherente con la

[10] Reproducimos el contenido literal del art. 140 del RD 1/2007 de 16 de noviembre por el que aprueba la Ley General para la Defensa de Consumidores y Usuarios y otras leyes complementarias.

Art. 140. 1: El productor no será responsable si prueba:

a) Que no había puesto en circulación el producto.

b) Que, dadas las circunstancias del caso, es posible presumir que el defecto no existía en el momento en que se puso en circulación el producto.

c) Que el producto no había sido fabricado para la venta o cualquier otra forma de distribución con finalidad económica, ni fabricado, importado, suministrado o distribuido en el marco de una actividad profesional o empresarial.

d) Que el defecto se debió a que el producto fue elaborado conforme a normas imperativas existentes.

e) Que el estado de los conocimientos científicos y técnicos existentes en el momento de la puesta en circulación no permitía apreciar la existencia del defecto.

Directiva 85/374/CEE, que trata de la responsabilidad por daños derivados de productos defectuosos (Díaz, 2018, 114).

Optar por el modelo de responsabilidad vicarial supondría reconocer cierta subjetividad, aunque mínima, a la acción automatizada. La posibilidad de una responsabilidad subjetiva de los robots me parece francamente exorbitante, aunque sería la más coherente si se reconociese una verdadera personalidad a los robots autónomos. Sólo encontraría justificación de producirse realmente el fenómeno de la singularidad, de otro modo, siendo el robot incapaz de culpabilidad e incapaz en el sentido estricto de autonomía moral, no podría obrar por culpa o negligencia, que constituyen el fundamento de la responsabilidad subjetiva (Moro, 2015, 538).

Incluso, de darse la hipótesis de la singularidad, reconocer personalidad completa a los robots generaría el doble peligro señalado por Louiseau (2018, 1041-1042) del igualitarismo y la lucha identitaria: a) reconocerle personalidad a los robot sería iniciar el camino hacia su igualación progresiva con el ser humano y acabarían disfrutando de los derechos y privilegios que se establecen para la protección y defensa de la dignidad humana; y b) una vez les fuera reconocida esta igualdad de derechos, comenzaría la lucha identitaria para defender las diferencias y la excelencia del robot respecto de los humanos y el consecuente deber para los hombres de respetar estas diferencias o reconocerles esta excelencia.

Un escenario en que el que se reconociese subjetividad jurídica a los robots inteligentes no dejaría de ser amenazante para el ser humano, pues, como se ha señalado en la doctrina de nuestro país, supondría reconocerles capacidad jurídica para obligarse, ser titular y ejercer derechos, siendo de todo punto imprevisibles en su forma de actuar y pudiendo originar gravísimos daños. Por todo ello es siempre preferible mantener la consideración de cosas a los sistemas de Inteligencia Artificial fuerte, ya que se adapta además mejor a sus características (Núñez, 2018, 25).

Cuestión diferente sería discutir acerca de la necesidad de proteger a los robots autónomos, pero para esto no es necesario considerarlos personas electrónicas o ciber-físicas. Las cosas también pueden estar protegidas por el Derecho. Debemos considerar que lo que conforma una sociedad no son los derechos, sino las obligaciones recíprocas ordenadas al bien común y este bien lo es tanto de los hombres como de las cosas, sin que sea necesario reconocerles derechos a estas últimas. Bastaría con proteger a los robots del capricho o de los excesos de los hombres, mediante el establecimiento de unos deberes precisos en la proyección, fabricación, uso y mantenimiento de estas nuevas figuras tecnológicas (Medina, 2013b, 752). Así mismo, sería preciso, establecer un sistema de responsabilidad adecuado para cubrir los daños y eventuales

perjuicios que pudieran derivar del defectuoso cumplimiento o incumplimiento de estos deberes.

Dicho de un modo más sencillo, muchos de los problemas teóricos y prácticos que plantea la personalidad jurídica de los robots derivan de abordar esta cuestión desde el prisma de los derechos subjetivos, descuidando la esfera de los deberes que son consustanciales a la convivencia social y a la misma condición ciudadana (Greco, 2010, 339). En este ámbito, lo importante son los deberes y éstos, por el momento, son una realidad exclusivamente humana

6. CONCLUSIÓN

En las páginas precedentes se ha intentado reflejar algunas de las preocupaciones y polémicas que subyacen en torno a la personalidad electrónica de los robots, desde una visión antropomórfica de estos sistemas de IA en clave religiosa, filosófica, neurológica y jurídica. A lo largo de este trabajo se ha tratado de mostrar el error de atribuir cualidades estrictamente humanas a los robots autónomos, puesto que hoy por hoy se trata más de una cuestión meramente especulativa que práctica; en tanto que los sistemas de IA, incluso los más avanzados, distan mucho de poder emular o asemejar la potencialidad de la inteligencia humana.

Desde una perspectiva religiosa parece claro que debe rechazarse la personificación de los robots en la medida en que carecen de la dignidad ontológica del ser humano como creatura de Dios a su imagen y semejanza. Ciertamente en el Talmud se recogen noticias de seres vivos artificiales, el golem; pero este es visto como una deformación grotesca del hombre y del acto de creación, que tiende a desembocar en situaciones desastrosas e imprevisibles. La imagen del golem bien pudiera servir para representar el paradigma de los robots autónomos, incluso la llamada IA fuerte, y, del mismo modo, se tratarían de entidades grotescas, una deformación del ser humano -reducido sólo a una operatividad cuantitativa- cuyas consecuencias podrían ser terriblemente amenazadoras para la humanidad.

En un sentido filosófico, se ha cuestionado críticamente la verdadera naturaleza de la autonomía del robot inteligente y parece que está muy lejos de la libre determinación kantiana o de la sustancia racional tomista, ambas cualidades ontológicas por las que se conoce al hombre. En consecuencia, no se comprende razón alguna para considerar como agentes morales a estos sistemas de IA, en la medida en que están determinados por su programación a actuar sin que en estos procesos jueguen papel alguno el sentido del deber, la culpa o la propia responsabilidad. Cuestión diferente es su uso como instrumentos para la toma de decisiones morales; pero, en todo caso, esta opción presenta unos riesgos que no deben ser minimizados y debería ofrecerse

una formación ética a los usuarios que evitase el peligro de la aparición de automatismos en los procesos de decisión.

Desde la óptica de la neurociencia se ha buscado mostrar las insuficiencias de la visión empirista o behaviorista del cerebro humano, asumiendo la tesis de que el cerebro humano es una realidad biológica enormemente compleja que no entendemos del todo y, por ello, no es posible reproducir tecnológicamente. Los intentos en esta línea serán siempre fuente de frustración, porque la realidad biológica es muy diferente de la realidad mecánica o computacional y por más que se trate de emular no podrá alcanzar la complejidad bilógica de los seres vivos. Si se acepta la posibilidad de una teoría evolutiva de la IA, debe matizarse que se tratará de una evolución en clave mecánica o computacional, no biológica; por lo que no parece pueda una máquina alcanzar la consciencia, la eticidad y emotividad propia del ser humano. Todo intento de introducir en una máquina unos patrones formalizados o un código de programación en clave ética y emocional -una conciencia- estará siempre abonado al fracaso en tanto se tratará de una formalización ridícula y banal de la realidad existencial y de la unicidad del ser humano.

En un sentido estrictamente jurídico, tras analizar diferentes posiciones, se ha defendido el estatus jurídico de los robots autónomos y sistemas de IA como cosas u objetos, lo que no debe implicar en ningún caso una desprotección de los mismos; sino realzar la función que cumple el deber en toda relación humana y jurídica y su trascendencia para el bien de la comunidad. De modo que lo crucial, mas que reconocer derechos o personalidad a estos robots, será establecer un sistema de deberes en el uso de estos y sistemas de IA que evite el abuso y la desviación en su manejo y disfrute. Está claro, el sujeto del deber es y seguirá siendo el hombre, no la máquina.

Finalmente, por remarcar la importancia del deber en este escenario, coincido con Wiener (1962) al afirmar que la ciencia ha de contribuir a la estabilidad y bienestar del conjunto social, que todo científico es un agente social y tiene inexorablemente una responsabilidad social con el bien común. Así, frente a las perturbaciones y amenazas que pueda presentar la incertidumbre del futuro, debe cuidar del jardín, que es la ciencia, para que crezca correctamente dando los mejores frutos. En este sentido, los avances científicos y tecnológicos de nuestra era deben servir al hombre y al bien de la humanidad en su conjunto, no al bien de unos pocos ni a sus ambiciones de dominio.

7. BIBLIOGRAFÍA

Arenas Dolz, Francisco (2016), "Cognición y retórica: bases biológicas del significado y la comprensión", en: *Pensamiento,* 72, 273, 997-1018.

Asís Roig, Rafael (2018), "Robótica, Inteligencia Artificial y Derecho", en: *Revista de Privacidad y Derecho digital* 10, 27-77.

Asaro, Peter M. (2012), "A Body to Kick, but Still No Soul to Damn: Legal Perspectives on Robotics", en: Abney, K. *et al.* (eds), *Robot Ethics: The Ethical and Social Implications of Robotics,* MIT Press, Cambridge, 169-186.

Asimov, Isaac (1985), *El hombre del Bicentenario*, Orbis, Barcelona.

Bellver Capella, Vicente, (2013), "Similitud e incomparabilidad divinas en el contexto del debate en torno al primado de la existencia y de la quididad en la filosofía islámica tradía", en: *Anales del Seminario de Historia de la Filosofía* 30, 2, 55-80.

Ben-Naftali, Orna, Zvi Triger (2013), "The human Conditioning: International Law and Science-Fiction", en: *Law. Culture and Humanities* 14, 1, 6-44.

Borelle, Céline (2018), "Sortir du débat ontologique: Éléments pour une sociologie pragmatique des interactions entre humáins et etres artificialles", en: *Reseaux* 212, 207-231.

Broman, Morgam, Pamela Fickenbercg-Broman (2018), "Socio-Economics and Legal Impact of Robotics and Autonomous Robotics and AI Entities. The RAiLE Project", en: *IEEE Techonology and Society Magazine*, Marc, 70-79.

Capellini, Alberto (2019), "Machina delinquere non potest? Brevi appunti sul Intelligenza Artificiale e responsabiltá penale", en: *Criminalia*¸ 2018, 499-520.

Chalmers, David J. (2010), "The Singularuty: An Philosophical Analysis", en: *Journal of Consciousness Studies* 17, 7-65.

Corey Jr, Gerald. A (2000), "From Maclean's triune brain concept to the conflict systems neurobehavioral model: the subjective basis of moral and spiritual consciousness", en: *Zygon*, 35, 2, 385-414.

De Castro y Bravo, Federico (1949), "La Sociedad Anónima y la deformación del concepto jurídico de persona", en: *Anuario de Derecho Civil*, 2, 4, 1397-1418.

Daley, Mark (2018), "Lo que la evolución nos puede enseñar acerca de la inteligencia artificial", en: *Revista de Occidente*, 446-447, 35-47.

De Aquino, Tomás, (2001), *Suma teológica,* I, BAC, Madrid.

Díaz Alabart, Silvia (2018), *Robots y responsabilidad civil,* Reus, Madrid.

Echevarría, Martín Federico (2019), "La mente como imago Dei según Tomás de Aquino", en: *Espíritu,* LXVIII, 223-252.

Ercilla, Javier (2018a), "Aproximación a una Personalidad Jurídica Específica para los robots", en: *Revista Aranzadi de Derecho y Nuevas Tecnologías,* 47.

— (2018b), *Normas de Derecho Civil y Robótica*, Aranzadi, Cizur Menor.

Frison-Roche, Marie-Anne (2017), "La disparition de la distinction de jure entre la personne et les choses: gain fabuleux, gain catastrophique", en: *Recueil Dalloz*, 2386-2389.

Fuselli, Stefano (2018), "*Logoi enuloi*. Aristotle's Contribution to the Contemporary Debate on Emotions and Decision-Making", en: Huppes-Cluysenaer, L., N. Coelho (eds.), *Aristotle on Emotions in Law and Politics*. Law and Philosophy Library, vol 121. Springer, Cham, 91-111.

García Sánchez, María Dolores (2020), "Inteligencia artificial y oportunidad de creación de una personalidad electrónica", en: *Revista Ius et Scientia*, 6, 2, 83-95.

Greco, Tommaso (2010), "Antes el deber: una crítica de la filosofía de los derechos", en: *Anuario de Filosofía del Derecho,* XXVI, 327-344.

Hegel, Georg Friedrich Wilhelm (1993), *Fundamentos de la Filosofía del Derecho,* Ensayo, Madrid.

Herrera de las Heras, Ramón (2022), *Aspectos Legales de la Inteligencia Artificial. Personalidad jurídica de los robots, protección de datos y responsabilidad civil*, Dykinson, Madrid.

Hubbard, F. Patrick (2011), "Do androids dream? Personhood and intelligence artificial", en: *Temple Law Review,* 83, 405-474.

Kant, Immanuel (2008), *Metafísica de las Costumbres*, 4ª ed, Tecnos, Madrid.

Kelsen, Hans (2011), *Teoría Pura del Derecho. Primera edición 1934,* Trotta, Madrid.

Lacruz Mantecón, Miguel L. (2020), *Robots y personas. Una aproximación jurídica a la subjetividad cibernética*, Reus, Madrid.

Lahabi, Mohamed Aziz (1967), *Le personnalisme musulmán*, 2ª ed, Puf, Paris.

Larregue, Julien (2019), "Un tournant relativiste chez les juristes ? La distinction entre les personnes et les choses n'est pas menacée par les robots humanoides", en: *Carnet Zilsel*, nº 5 (disponible en zilsel.hypotheses.org).

Loiseau, Grégoire (2018), "La personalité juridique des robots: une monstruosité juridique", en: *La semaine juridique* 22, 1039-1042.

MacLean, Paul D. (1990), *The Triune Brain in Evolution: Role in Paleocerebral Functions*, Springer, New York.

Massignon, Louis (1952), "Le respect de la personne humaine en islam, et la priorite du droit d'asile sur le devoir de juste guerre", en: *Revue International de la Croix-Rouge,* 6, 448-468.

McFee, Graham (2019), *Philosophy and the 'Dazzling Ideal' of Science*, Palgrave, Macmillan, Switzerland.

Medina Morales, Diego (2013a), "Sujeto o Persona, de la sustantividad a la formalidad de un concepto", en: López Hernández, José, Fernando Navarro Aznar, José Ramón Torres Ruiz (dirs.), *Estudios de Filosofía del Derecho y Filosofía Política Homenaje al Profesor Alberto Montoro Ballesteros*, Ediciones de la Universidad de Murcia, 623-633.

— (2013b), "Ius bestiarum, más allá del ius Gentium"; en AAVV, *Studi in onore di Augusto Sinagra*, Aracne, Milan, 741-757.

Moro, Paolo (2015), "Libertà del robot? Sull'etica delle machine intelligenti", AAVV, *Filosofía del Diritto e Nueove Tecnologie. Prospettive di ricerca tra teoria e pratica*, Aracne, Aricia, 1-20.

Morote Sarrión, Juan (2009), *El concepto de persona. Una aproximación interdisciplinar*, Obra abierta, Valencia.

Nagenborg, Michael *et al.* (2008), "Ethical regulations on robotics in Europe", en: *AI and Society*, 22, 3, 349-366.

Nietzsche, Friedrich (1965), *Obras Completas. Así habló Zaratustra*, II, Aguilar, Buenos Aires.

Nuñez Zorrila, Mª Carmen (2018), "Los nuevos retos de la Unión Europea en la regulación de la responsabilidad civil por los daños causados por la Inteligencia Artificial", en: *Revista Española de Derecho Europeo,* 66, 9-53.

Polo, Leonardo (1997), *Nominalismo, idealismo y realismo*, Eunsa, Pamplona.

Popper, Karl, John C. Eccles (1985), *El yo y su cerebro*, Labor, Barcelona.

Rodríguez Puerto, Manuel (2021), "¿Puede la Inteligencia Artificial interpretar normas jurídicas? Un problema de razón práctica", en: *Cuadernos Electrónicos de Filosofía del Derecho,* 44, 74-96.

Savulescu, Julian, Hannah Maslen (2015), "Moral enhancement and artificial intelligence: moral ai?", en: Romport, J. *et al.* (eds.), *Beyond Artificial Intelligence*, Springer, Switzerland, 79-95.

Searle, John, (2000), *El misterio de la conciencia*, Paidós, Barcelona.

Sünderhauf, Niko *et al.* (2018), "The limits and potential of Deep learning for robotics", en: *The Internartional Journal of Robotics Research* 37, 4-5, 405-420.

Trachtenberg, Joshua (2004), *Jewish Magic and Superstition: A Study in Folk Religion*, University of Pensilvania Press, Philadelphia.

Wiener, Norbert (1962), "Science and society", en: *Science* 138, n. 3541.

— (1966), *God and Golem, Inc,* Massachusetts Institute of Technologie, Cambridge.

Wilcocks, Leslie (2020), "Robo-Apocalipse canceled? Reframig the automation and future of work debate", en: *Journal of Information Technologie* 35, 4, 286-302.

Yue-Hsuan, Weng etl. (2009), "Toward the Human-Robot co-existence society: on safety intelligence for next generation robots", en: *International Journal Social Robotics* 1, 267-282.

Zimmerman, Evan J. (2015), "Machine minds. Frontiers in Legal Pesonhood", en: *Social Science Research Network Electronic Journal*, 41, 1-43.

CAPÍTULO XV

DERECHO DEL TRABAJO, INTELIGENCIA ARTIFICIAL Y ROBÓTICA[1]

MARÍA SEPÚLVEDA GÓMEZ
Universidad de Sevilla
mariasep@us.es

"El trabajo no es una mercancía"
(OIT. DECLARACIÓN DE FILADELFIA, 1944)

1. LA TRANSFORMACIÓN DIGITAL. UN NUEVO PARADIGMA TÉCNICO-ECONÓMICO Y SOCIAL

Se afirma que la transición a la primera fase del capitalismo de alta tecnología bajo el signo de la automatización se produjo con el traslado del uso de la computadora, originada en el ámbito militar, al ámbito de producción económica y también a la producción de conocimientos en el terreno de las ciencias experimentales, en la segunda mitad del siglo XX, y esto hizo época (Haug, 2016, 23).

Más adelante, es la aparición gradual, desde principios de los años 70 del siglo pasado, de un conjunto de tecnologías de la información y la comunicación que permiten la hibridación entre el mundo físico y el digital, borrando las fronteras entre ambos, lo que ha llevado a hablar de una cuarta revolución industrial (Braña Pino, 2020). También denominada "Industria 4.0", impulsada por la transformación digital, ha supuesto un salto cualitativo en la organización y gestión de la cadena de valor del sector (Ministerio de Industria, Energía y Turismo, 2015).

El uso de la inteligencia artificial (IA), considerada como tecnología de las tecnologías, va a ser el elemento fundamental de esta nueva industria 4.0, y se ha definido como un marco que engloba diferentes especialidades que comparten un objetivo común: dotar a un sistema artificial de cierto grado de inteligencia (Del Rey, 2018, 32), con capacidad de producir unos resultados de razonamiento equivalentes a los obtenidos por la inteligencia natural humana.

[1] Este trabajo se ha realizado en el del marco del Proyecto I+D *Política de rentas salariales: salario mínimo y negociación colectiva*, ref. P20_01180. Un primer avance de este trabajo fue expuesto en la Segunda mesa redonda: "Inteligencia Artificial y Robótica en el marco del Estado de Derecho" del Congreso Internacional sobre Inteligencia Artificial, Robótica y Filosofía del Derecho, organizado por: Prof. Dr. Fernando H. Llano Alonso, Prof. Dr. Álvaro Sánchez Bravo, Prof. Dr. César Villegas Delgado (codirectores), celebrado en la Facultad de Derecho de la Universidad de Sevilla, en diciembre de 2021.

De todas las posibles, el sistema de IA con capacidad de aprendizaje es la que puede tener mayor impacto en el ámbito laboral, a través de una red neuronal artificial y de algoritmos autodidácticos con capacidad de ampliar sus conocimientos y toma de decisiones (Del Rey, 2018, 191-192).

En efecto, en esta denominada nueva era digital, o cuarta revolución industrial, el actual proceso de transformación digital va mucho más allá del mero uso de diferentes tecnologías. Se caracteriza por la combinación de diversas tecnologías con capacidad de producir nuevas tecnologías disruptivas, que alteran el *statu quo* establecido en un determinado ámbito para crear un nuevo y diferente (Mercader, 2017, 30), pero no sólo trae cambios tecnológicos, sino también una transformación fundamental de nuestra sociedad (Widuckel/Aschenbrenner, 2020).

En este sentido, se afirma que lo que viene ahora es la "automatización automatizada", que se basa en los progresos de la llamada inteligencia artificial y del hardware y el software del proceso de datos a gran escala en conexión con un conjunto de tecnologías, ante todo los sensores, que permiten salvar el abismo entre la existencia virtual y la real. Esto es algo más que una ampliación lógica de las técnicas anteriores, porque abre la puerta a modelos de negocio enteramente nuevos y posibilita la aparición de nuevos productos" (Haug, 2016, con cita de Giersberg).

En efecto, en esta nueva era el término digitalización no significa (sólo) que las empresas se digitalizan, sino que transforman sus procesos, de manera que surgen nuevos productos y servicios, con el análisis de macrodatos que averiguan las preferencias de los consumidores y usuarios, y nuevas formas de trabajo a través de procesos productivos digitalizados, que sustituyen fases, tareas y actividades por otras nuevas y diferentes o, simplemente desaparecen porque ya no son necesarias. El caso más paradigmático de nueva forma de empleo y de ruptura con la tradicional relación laboral ha sido el del trabajo a través de plataformas digitales, pero no es el único. O la implantación monetarista de sistemas digitales que buscan la competitividad a costa de la reducción de puestos de trabajo y, por tanto, la reducción de costes y una mayor obtención de beneficios.

El sector de la banca ha sido uno de los pioneros en la digitalización, probablemente por la globalización de los mercados financieros y la competitividad a nivel global del sector. Pero, mientras en unos países ha supuesto literalmente la desaparición de los puestos de trabajo ahora digitalizados, en otros, no ha sido (tanto) así. A título ilustrativo, en España, en el periodo 2008-2015 redujo un 32,5% su número de oficinas y un 28,9% el número de empleados; en Alemania, en cambio la proporción fue de 13,9% y 5,7% respectivamente. Se constata que España, junto a Grecia e Irlanda son los

países donde el ajuste del empleo en la banca ha sido mayor en dicho periodo, siendo la caída reducida en países como Alemania, Francia o Italia. (Cruz-García/Maudos, 2016, 87).

El efecto de destrucción de empleo en la industria y servicios por la implantación de procesos de IA y robotización está siendo, no obstante, objeto de las políticas públicas de empleo, tanto a nivel nacional como europeo. La UE pronostica que la doble transición ecológica y digital podrá crear para 2050 hasta dos millones de puestos de trabajo adicionales en toda la Unión, aunque a corto plazo produzca desempleo, y los empleos de ciertos sectores y regiones correrán el riesgo de ser desplazados[2].

La importancia de este proceso transformador trasciende el plano de la producción de bienes y servicios para ir más allá. Para Haug (2016), el llamado Internet de las cosas y sus aplicaciones específicas a los procesos industriales de transformación, bajo el rótulo de Industria 4.0, se ha situado en un primerísimo plano de la atención económica y política. Para el autor, el motor de desarrollo de la tecnología más avanzada es la obtención de beneficio o, dicho de otro modo, la actualización de la tecnología depende del beneficio que se espera obtener de ello. Por tanto, ese factor es donde se localiza la capacidad o el poder para configurar y hacer operativas las iniciativas, de realizar las posibilidades. Su modo de funcionamiento es la competencia, sus actores primarios son los capitales en concurrencia, y los secundarios, los estados. El medio de realización de unos y otros es el ámbito de las necesidades de la población. De este modo, el complejo tecnocientífico, que responde al mito de la Industria 4.0, se ha traducido en la tecnología de la distancia, es decir, una tecnología cuya función principal consiste en mantener a distancia a los competidores, pero el impulso del desarrollo de las tecnologías para la consecución de la competitividad de los estados y del capital, sirve para exportar desempleo (Haug, 2016, 25-27).

Se afirma que la UE tiene unos objetivos y estrategias geopolíticas en una competición con otras potencias económicas por la soberanía digital, que se enmarca en el concepto más amplio de autonomía estratégica, y que ha cobrado fuerza ante la necesidad de acelerar la recuperación económica tras la pandemia, reducir el desfase industrial y tecnológico europeo para competir por los mercados mundiales, enfrentar las presiones geopolíticas en relación al 5G o la ocupación de tramos de la cadena de valor digital por parte de terceros. Pero también que pretende contribuir a la configuración de reglas y normas mundiales sobre la base de los valores europeos, los derechos fundamentales, la

[2] *Recomendación 2021/402, de 4 de marzo de 2021, sobre un apoyo activo eficaz para el empleo tras la crisis de la COVID-19 (EASE).*

seguridad y la garantía del modelo económico y social europeo" (Consejo Económico y Social España, 2021, 37-38).

Todavía está por ver cuál será el modo y el resultado de esta espiral por la soberanía digital en el marco de un capitalismo tecnológico. Lo que se denomina como una nueva revolución industrial es la transición hacia un nuevo paradigma técnico-económico, basado en las tecnologías de la información y las redes, el desarrollo de la economía basada en el conocimiento, el cambio en la provisión de necesidades colectivas y una reconfiguración de las relaciones sociales. Un nuevo paradigma donde la materia prima es la información. Es el paso de un capitalismo industrial a un capitalismo digital, que ya en pleno siglo XXI incorpora la inteligencia artificial (Braña Pino, 2020).

Pero la digitalización es (también) un proceso social en construcción, que requiere la articulación de una estrategia integral para la transición justa a la economía digital, que favorezca la creación de empleo decente, y contribuya en paralelo a prevenir y mitigar los riesgos de segmentación y exclusión social entre la población (Rocha, 2019, 10).

Esa necesaria estrategia integral para que la transición digital sea justa, es en la actualidad el reto de los ordenamientos jurídico-laborales, las políticas públicas, y la negociación colectiva. El nuevo paradigma técnico-económico es también social. El trabajo inteligente, flexible y eficaz no justifica una ruptura del equilibrio de intereses en la relación de trabajo, ni una completa desprotección social.

Una transición justa, que requiere de una ordenación jurídica de la relación de trabajo que no debe cuestionar sino, al contrario, fortalecer en ese nuevo entorno muchos de los valores y principios más propios del Estado social, eludiendo o superando muchas reglas al solo servicio de la lógica económica, de dudosa sostenibilidad incluso en términos estrictamente democráticos (Valdeolivas, 2022, 191).

2. GRADO DE DIGITALIZACIÓN DEL TEJIDO EMPRESARIAL ESPAÑOL

Pese a que se constata que el proceso de transformación digital se caracteriza por la velocidad con la que experimenta cambios, no obstante, su implantación y desarrollo puede ser muy dispar, no sólo entre territorios, sino también entre sectores productivos. No podemos abordar aquí los factores que influyen en esto, pero sí nos parece de interés ofrecer algunos datos aproximativos sobre el grado de digitalización de las empresas españolas -especialmente en IA y robótica- según estadísticas nacionales, y algunos datos de la UE al respecto, lo que nos permitirá contextualizar mejor el alcance de la digitalización del trabajo en nuestro país.

La Agenda Digital 2025[3], presentada por el Gobierno el 23 de julio de 2020, reconoce que el progreso de la transformación digital en España ha sido más limitado en el terreno de la digitalización de la industria y la empresa -especialmente PYMEs-.

A pesar de ello, las estadísticas 2021 recogen un incremento en el uso de las tecnologías digitales por las empresas españolas, especialmente las grandes empresas. Según la Encuesta sobre el uso de TIC y del comercio electrónico en las empresas para el año 2020-2021 del INE[4], las tecnologías más utilizadas por las empresas españolas de más de diez trabajadores son las herramientas para gestionar la información dentro de la empresa (ERP = Enterprise Resource Planning), y las que gestionan la información de clientes (CRM = Customer Relationship Management), con un 51,7% y 41,8% de empresas, respectivamente. Ambas se han incrementado en 6,3 puntos respecto a la anterior edición. En cuanto al Internet de las Cosas (IoT) es la tecnología cuyo uso más se ha incrementado (10,9 puntos), siendo utilizado ahora por el 27,7% de las empresas. Por su parte, la utilización de análisis de Big Data alcanzó un 11,1% en el periodo 2020-2021, lo que supuso un incremento de 2,6 puntos más que en el periodo anterior 2019-2020. También experimenta un incremento el porcentaje de empresas que compran servicios en la nube (4,2 puntos), que asciende a un 32,4%.

Respecto de la tecnología de la Inteligencia Artificial (IA) es utilizada por el 8,3% de las empresas (primer trimestre de 2021), siendo en las grandes empresas donde se concentra el mayor porcentaje del uso de IA (empresa entre 10 y 49 empleados: 6,69%; entre 50 y 249 empleados: 13,57%; y de 250 empleados en adelante: 33,06%). En el caso de empresas con menos de 10 empleados, utilizan algún sistema de IA sólo el 3,47% del total[5].

Nos parece ilustrativo recoger el desglose de porcentajes de empresas españolas con diez o más empleado (primer trimestre 2021) que usan diversas tecnologías de IA, para diferentes fines, en las industria, construcción y servicios.

---

[3] https://portal.mineco.gob.es/RecursosArticulo/mineco/prensa/ficheros/noticias/2018/Agenda_Digital_2025.pdf

[4] INE, Encuesta sobre el uso de TIC y comercio electrónico en las empresas 2021-primer trimestre 2021. (Nota de prensa actualizada a noviembre de 2021). Consultable en https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176743&menu=ultiDatos&idp=1254735576692

[5] Fuente: INE. https://www.ine.es/dynt3/inebase/index.htm?padre=8287&capsel=8287.

Empresas con 10 o más empleados	Total %	Industria	Construcción	Servicios
H.1 % empresas que emplean tecnologías de Inteligencia Artificial (IA)	8,32	7,52	3,77	9,79
H.1.A % empresas con tecnología IA de análisis del lenguaje escrito	29,75	13,57	36,88	34,37
H.1.B % empresas con tecnología IA que convierte el lenguaje hablado en formato legible por una máquina	31,70	24,21	37,32	33,61
H.1.C % empresas con tecnología IA que genera lenguaje escrito o hablado	19,14	11,07	15,66	22,11
H.1.D % empresas con tecnología IA de identificación de objetos o personas en función de imágenes	40,56	43,65	29,69	40,58
H.1.E % empresas con tecnología IA de análisis de datos (Aprendizaje automático)	30,42	23,42	12,69	34,41
H.1.F % empresas con tecnología IA de automatización de flujos de trabajo o ayuda en la toma de decisiones	38,57	41,38	22,86	39,14
H.1.G % empresas con tecnología IA que permite el movimiento físico de máquinas	12,64	25,13	5,03	9,27
H.2.A % empresas que emplean IA para Marketing o ventas	22,18	9,46	12,72	27,26
H.2.B % empresas que emplean IA para procesos de producción	23,82	41,40	4,64	19,89
H.2.C % empresas que emplean IA para organización de procesos de administración de empresas	20,22	12,50	19,44	22,83
H.2.D % empresas que emplean IA para gestión de empresas	15,00	9,81	3,40	17,81
H.2.E % empresas que emplean IA para logística	10,76	14,92	4,64	9,97
H.2.F % empresas que emplean IA para seguridad de las TIC	21,80	24,18	11,48	22,01
H.2.G % empresas que emplean IA para gestión de recursos humanos o contratación	7,67	4,72	0,23	9,34
H.3.A % empresas cuya IA fue desarrollada por empleados propios	25,48	16,08	16,72	29,40
H.3.B % empresas con paquetes de IA comerciales modificados por empleados propios	17,69	9,30	15,51	20,64
H.3.C % empresas con paquetes de IA de código abierto modificados por empleados propios	16,37	10,99	3,11	19,39
H.3.D % empresas que compraron paquetes IA comerciales listos para usar	37,30	43,98	33,64	35,47
H.3.E % empresas que contrataron a proveedores externos para desarrollar/modificar los sistemas de IA	43,12	47,25	41,70	41,91

(Fuente: INE, Encuesta de uso de TIC y comercio electrónico en empresas 2020-2021. https://www.ine.es/index.htm)

Por comunidades autónomas, el porcentaje más elevado de empresas de diez o más empleados que usan alguna tecnología de IA, lo tiene la Comunidad Autónoma de Madrid (11%), seguida de Cantabria (9,56%), Castilla y León (9%), País Vasco (9%), y Canarias (8,33%).

En cuanto a la implantación de robótica[6], el porcentaje de utilización de algún tipo de robots por las empresas españolas de diez o más empleados, en 2020, es relativamente bajo, ya que alcanza sólo el 8,89%, y en el caso de empresas de menos de diez trabajadores supone tan sólo el 1,79% -la media en el ámbito de la UE se situó en 2020 en un 28% en grandes empresas, y un 6% en pequeñas y medianas empresas[7]-.

Del total de empresas con diez o más empleados que usan algún tipo de robot (8,89%)[8], el 77,23% utilizan robots industriales, especialmente las medianas empresas, y el 37,96% utilizan robots de servicio, especialmente las pequeñas empresas. El porcentaje más alto de empresa utilizan el uso de robots de servicio para tareas de limpieza o eliminación de residuos (44,22%), seguido de gestión de almacén (38,92%), vigilancia o seguridad (18,77%), transporte de personas o bienes (15,99%), trabajos de ensamblaje (13,59%), tareas de reparación de daños (10,90%), y finalmente tareas de dependiente de tienda robótico (4,3%).

Por otro lado, en el ámbito de la UE España ocupa el puesto número 9 de los veintisiete Estados miembros de la UE en el Índice de la Economía y la Sociedad Digitales (DESI), publicado por la Comisión Europea (edición 2021)[9]. Este Informe global mide cuatro indicadores: capital humano; conectividad; integración de la tecnología digital por las empresas españolas; y servicios públicos digitales. En el conjunto de estos indicadores, España tiene una puntuación de 57,4, siendo la media de la UE de 50,7.

Ahora bien, en el Informe DESI por países de la UE[10], España ocupa el puesto número 16 en el respecto de la integración de tecnología digital por las empresas españolas, coincidiendo su puntuación en este indicador con la media europea. El porcentaje de empresas española que utilizan servicios en la nube es del 22% (en comparación con la media de la UE del 26%). Por lo que hace al uso de IA, el porcentaje de empresas españolas es del 22%, siendo la media de la UE

[6] INE, Encuesta de uso de TIC y Comercio Electrónico (CE) en las empresas 2019-2020, consultado en: https://www.ine.es/dynt3/inebase/es/index.htm?padre=6880

[7] https://digital-strategy.ec.europa.eu/en/library/digital-economy-and-society-index-desi-2021

[8] INE, Encuesta de uso de TIC y Comercio Electrónico (CE) en las empresas 2019-2020, consultado en: https://www.ine.es/dynt3/inebase/es/index.htm?padre=6880

[9] https://digital-strategy.ec.europa.eu/en/policies/desi

[10] https://digital-strategy.ec.europa.eu/en/policies/desi

del 25%, y solo el 9% de las empresas accede a análisis de macrodatos (Big Data), en tanto que la media de UE se sitúa en el 14%. Además, se constata que estas tecnologías avanzadas son utilizadas principalmente por las grandes empresas, en tanto que un escaso porcentaje de pequeñas y medianas empresas la usan[11].

De los datos estadísticos apuntados, tanto nacionales como europeos, respecto de la utilización por las empresas española de tecnologías como IoT, servicios en la nube, Big Data, IA, y robótica, podemos extraer algunas conclusiones a grandes rasgos.

La primera, es el incremento generalizado en el periodo de referencia de empresas que usan tecnologías como IoT, Big Data, servicios en la nube, siendo la IoT la que más aumento de uso ha experimentado (10,9 puntos). Respecto de la IA y robótica llama la atención que, en comparación con el porcentaje de empresas que usan las otras tecnologías antes indicadas, en el caso de la IA y la robótica el número de empresas que la usan es bastante reducido y, además no existen datos de uso en periodos anteriores a 2020-2021.

En segundo lugar, y referido a la tecnología IA, cabría diferenciar tres ítems diferentes:

1) Según tipo de IA utilizada: se observa que la más utilizada por las empresas con diferencia es la IA que permite la identificación de objetos o personas en función de imágenes (industria y servicios), seguida de la IA de automatización de flujos de trabajo o ayuda en la toma de decisiones (industria y servicios), la IA que convierte el lenguaje hablado en formato legible por una máquina (construcción y servicios), empresas con tecnología IA de análisis de datos -aprendizaje automático- (servicios e industria), y empresas con tecnología IA de análisis del lenguaje escrito (construcción y servicios).

2) En función de los fines en el uso de la IA: de mayor a menor uso, destaca en especial el porcentaje de empresas que emplean IA para procesos de producción (industria), seguidas de las que la usan con fines de seguridad (industria y servicios), para la organización de procesos de administración de empresas (servicios e industria), para la gestión de la empresa (servicios e industria), para logística (industria y servicios), y finalmente con fines de gestión de los recursos humanos y contratación (servicios e industria).

3) En cuanto a la gestión de la tecnología IA: destaca con diferencia el número de empresas que contrataron a proveedores externos para desarrollar/modificar los sistemas de IA (industria y servicios), seguidas de las

[11] Índice de Intensidad Digital (IID) 2021. https://digital-strategy.ec.europa.eu/en/library/digital-economy-and-society-index-desi-2021

que compraron paquetes IA comerciales listos para usar (industria y servicios), y en menor medida las empresas cuya IA fue desarrollada por empleados propios, sobre todo en el sector servicios.

En los tres ítems expuestos se ven afectados diferentes aspectos de la relación de trabajo por el uso de tecnologías de IA: en el primer ítem, claramente a través de la identificación de personas por imágenes, lo que no resulta nada novedoso en el ámbito laboral, pues es algo que lleva años empleándose por las empresas como medios de control digital de la ejecución del trabajo, que involucra derechos fundamentales como la intimidad, la propia imagen, la protección de datos personales, entre otros.

El segundo ítem indicado, esto es, la finalidad para la que usan las empresas la IA, destaca su empleo en los procesos de producción, especialmente en la industria y en menor medida en el sector servicios. Sin duda es uno de los efectos más preocupantes del uso de IA, pues afecta directamente a la desaparición de puestos de trabajo, aunque esta afirmación es matizable.

El tercer ítem, referido al desarrollo y gestión de la propia tecnología IA, entronca directamente con la necesidad de transparencia e intervención humana en la aplicación de la IA. Es un aspecto que interesa tanto a las personas trabajadoras como a sus representantes, y su participación en la implantación y control de la IA, por los derechos que pueden verse afectados.

En tercer lugar, en relación al uso de la robótica por las empresas españolas se observa un porcentaje bajo de empresas con más de diez empleados que usan robots (8,89%), en relación con el de la media europea (28%). Y también un número muy reducido de empresas de menos de diez trabajadores que usan robots (1,79%) en relación con la media UE (6%). En cuanto al tipo de robots, destaca sobre todo el uso robots de industria, y en menor medida el uso de robots de servicio.

Es evidente que el uso de robots en la industria ha afectado, y lo sigue haciendo, a la reducción de puestos de trabajo, de manera que es una tecnología que afecta directamente al empleo, que requiere entre otras actuaciones definir mecanismos de identificación de nuevas capacitaciones (Valverde Asencio. 2019:141), sin olvidar la implicación que tiene también en el ámbito de la seguridad y salud de las personas trabajadoras en los casos de robots colaborativos, ya que la tecnología empleada en el trabajo genera riesgos muy diversos, incluidos los psicosociales (Purcalla Bonilla, 2020, 114).

En cuarto lugar, tanto respecto de la IA como de la robótica se ha constatado su mayor empleo en grandes empresas (en el caso de la IA, un 33% del total de empresas que usan IA, emplean a 250 trabajadores o más). También

se ha constatado el ínfimo uso que hacen las pequeñas empresas de estas tecnologías. Podría pensarse, por tanto, que en España estas tecnologías tienen poco impacto en el trabajo, ya que el 99% de las empresas son pequeñas y medianas. Pero esta afirmación necesita matización.

Según el MTES[12], a finales de septiembre de 2021 el número de empresas con trabajadores fue de 1.309.569, de las cuales algo menos de 5.000 son grandes empresas (las que emplean más de 250 trabajadores), sin embargo, del total de personas trabajadoras empleadas (14.248.998), un 40% trabaja para grandes empresas, es decir, cinco millones setecientas mil, aproximadamente. De manera que no es despreciable en absoluto el número de personas trabajadoras que ocupan puestos en los que se usan estas tecnologías, son seleccionados con el uso de procesos de IA, ejecutan trabajos controlados por dispositivos digitales, realizan tareas con robots colaborativos, o pierden su empleo por la digitalización de tareas y funciones.

Además, se espera un incremento en la digitalización de empresas, ya que la transformación digital de la economía y de las empresas es un proceso en movimiento, y es objetivo de la UE que a 2030 el 75% de las empresas alcancen su transformación digital mediante el uso de las tecnologías de IA, Big Data y servicios en la nube. La transformación digital es uno de los cuatro puntos cardinales de la denominada Brújula Digital de la UE[13].

En 2020, España adoptó una nueva y ambiciosa agenda digital, España Digital 2025, que incluye cerca de 50 medidas agrupadas en diez ejes estratégicos con los que se pretende impulsar el proceso de transformación digital del país durante los próximos cinco años, de forma alineada con la estrategia digital de la Unión Europea, mediante la colaboración público-privada y con la participación de todos los agentes económicos y sociales del país[14]. El Plan de Recuperación, Transformación y Resiliencia español, que cuenta con los nuevos instrumentos comunitarios de financiación *Next Generation EU* (la digitalización tiene que ser uno de los ejes principales para movilizar estos recursos), cuyo presupuesto total asciende a 69.500 millones de euros, contiene un conjunto ambicioso de reformas e inversiones en el ámbito digital. El Plan destina el 28,2% del total de los fondos invertidos al ámbito digital, y se centra especialmente en promover la

[12] https://www.mites.gob.es/estadisticas/emp/Emp21-Sep/Resumen%20de%20resultados%20Septiembre%202021.pdf

[13] https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_en.

[14] https://portal.mineco.gob.es/es-es/ministerio/estrategias/Paginas/00_Espana_Digital_2025.aspx.

digitalización de las empresas, especialmente de las pymes (a las que se destina el 25% del presupuesto total para el ámbito digital)[15].

Pero no cabe olvidar que también es objetivo del Plan de Recuperación, Transformación y Resiliencia "garantizar los derechos en el nuevo entorno digital, y en particular, los derechos laborales", entre otros.

3. INTELIGENCIA ARTIFICAL, ROBÓTICA Y RELACIONES DE TRABAJO

La preocupación por el impacto de la técnica ha sido siempre una constante (Mercader, 2017, 30), y los efectos de los grandes cambios no se muestran de forma inmediata, sino a medio plazo. Es lo que ha ocurrido en el pasado, desde la primera revolución industrial, con la introducción de la máquina y el uso de la energía de vapor, la posterior organización científica del trabajo y la producción masiva y en serie. A mediados del siglo XX, fue el auge de los dispositivos electrónicos, las tecnologías de la información, y la implantación de líneas automatizadas de producción, constituyendo un nuevo salto tecnológico, con sistemas de producción flexibles. Sus efectos llegan a principios del presente siglo XXI, como momento álgido de estas transformaciones: la sociedad postindustrial, del conocimiento y de la información (Mercader, 2017, 27-28), y una economía globalizada y digitalizada, con la expansión del uso de tecnologías avanzadas en todos los ámbitos de la vida y, en especial en el trabajo, tales como la IA y la robótica.

Si, como se ha afirmado, la informática ha sido el instrumento que ha hecho posible la globalización, que ha anulado el tiempo y el espacio, al menos en las actividades inmateriales (Losano, 2006, 20), en el momento actual todavía no se puede saber con seguridad cuál será el resultado en un futuro no muy lejano del actual complejo proceso de transformación digital que está experimentando la economía y la sociedad en su conjunto. Probablemente no podamos identificar un único resultado final, sino múltiples resultados que, en evolución constante, afectarán a los más diversos ámbitos de la vida, y en buena parte ya lo está haciendo. De ahí que se afirme que no estamos propiamente ante una nueva revolución, la digital, sino ante un proceso de revoluciones sucesivas, de mutaciones, de evoluciones, de rupturas, de reequilibrios en cadena (Mercader, 2017, 24, con cita de Braudel).

Todas estas transformaciones están impactando en el trabajo humano y su ordenación jurídica, cuyo concreto alcance cualitativo y cuantitativo no resulta nada pacífico (Cruz, 2017, 14). Transformaciones que se incorporan a un contexto de globalización de la economía y de un mercado financiero global, la

[15] https://www.lamoncloa.gob.es/temas/fondos-recuperacion/Documents/30042021-Plan_Recuperacion_%20Transformacion_%20Resiliencia.pdf.

deslocalización empresarial, la dispersión de la parte empleadora de la relación de trabajo, y de los puestos de trabajo, el auge del trabajo a distancia, los cambios demográficos, los nuevos flujos migratorios, entre otros cambios a nivel mundial. La conjunción de la digitalización en un contexto de globalización hace que sus efectos impacten exponencialmente sobre el devenir del empleo y las relaciones laborales (Cruz, 2017, 16).

Los efectos de la digitalización en el empleo y la relación laboral son múltiples, y la doctrina científica y judicial se encuentra en la actualidad ante numerosas dificultades de interpretación jurídica de instituciones jurídico-laborales, cuya regulación no ha sido concebida para un contexto social y económico en proceso de transformación digital. Pero esto ha sido siempre una característica del Derecho del Trabajo: la realidad social y económica que regula sale de los márgenes de la norma laboral, es cambiante, y se produce el desfase temporal entre norma y realidad, lo que no significa contraposición ideologizada entre lo *nuevo y lo viejo*, donde el Derecho del Trabajo aparece representado por su vetustez inservible (Trillo, 2016, 81). Lo más problemático en la actualidad es que esa realidad cambia a mucha velocidad, y la distancia entre ambas es mayor.

Desde esta perspectiva, la mayor parte de la doctrina científica coincide en la necesidad de una adaptación del ordenamiento jurídico-laboral a la nueva realidad digital. Es algo que se está haciendo de forma parcial, aunque progresiva, lo que no parece ser suficiente, por lo que se aboga por una reforma integral y sistémica de nuestro Estatuto de los Trabajadores y normas concordantes, y de la propia regulación de la Seguridad Social, reclamando una construcción renovada de nuestro modelo normativo e institucional de relaciones de trabajo en clave digital, adaptado y formulado sobre esa nueva realidad productiva, económica y de empleo que toca la práctica totalidad del modelo tradicional, por su afectación a casi cualquier tipo de trabajo, empresa y trabajador (Valdeolivas, 2022, 191-192).

Especial impacto está teniendo el uso de algoritmos en los procesos de selección para el empleo y contratación de personas. El análisis de macrodatos, a través de sistemas de aprendizaje automático, ha facilitado la creación de perfiles y permite la toma de decisiones automatizadas, (Sáez, 2020, 42-43). Estos sistemas de IA pueden contener sesgos que producen discriminaciones en el empleo, o en las condiciones de trabajo en función del sexo de la persona trabajadora, o su lugar de residencia, etnia, etc., pudiendo verse afectados derechos fundamentales como la intimidad, la protección de datos personales, y la igualdad y no discriminación (Sáez, 2020, 43). Y es que nos hallamos ante una situación de concurrencia directa entre el ser humano y la máquina, afectando

al corazón de los derechos fundamentales, esto es, a la dignidad de la persona trabajadora (Escande-Varniol, 2020, 152).

Una de las posibles soluciones que se proponen ante este tipo de impacto es una actualización de la tutela antidiscriminatoria frente a la discriminación algorítmica, considerada como categoría jurídica, mediante la introducción en nuestra normativa laboral de la obligación legal de la empresa de evaluar el impacto discriminatorio de las decisiones automatizadas, incluida la elaboración de perfiles (Sáez, 2020, 59). Lo que seguramente habrá de hacerse si se llega a aprobar la propuesta europea de Reglamento por el que se establecen normas armonizadas en materia de inteligencia artificial (ley de inteligencia artificial)[16], que considera como de alto riesgo los sistemas de IA "destinados a utilizarse para la contratación o selección de personas físicas, especialmente para anunciar puestos vacantes, clasificar y filtrar solicitudes o evaluar a candidatos en el transcurso de entrevistas o pruebas" (art. 6 en relación con el Anexo III), estableciendo entre otras garantías la obligatoriedad de aplicar un sistema de gestión de los riesgos de la IA.

De bastante calado ha sido también el impacto que ha tenido en el ámbito laboral el trabajo a través de plataformas digitales. Nueva forma de empleo que rompe con todos los esquemas del marco regulador laboral, y que ha dado lugar a una alta litigiosidad en torno a la calificación de la relación jurídica entre trabajador y plataforma, sobre todo en el caso de los repartidores a domicilio, conocidos como *riders*. A ello ha dedicado la doctrina laboralista gran cantidad de estudios en los últimos años, pues alrededor de la fórmula de contratación como trabajo autónomo han girado también otras cuestiones, como la calificación de la plataforma digital como empleador o mero intermediario, el uso de algoritmos para calcular la retribución en función de la valoración de los servicios por parte de los usuarios, o mantener o no la actividad en función del cumplimiento de objetivos fijados por la plataforma a través de algoritmos, entre otras.

Después de numerosas resoluciones judiciales contradictorias, el Tribunal Supremo declaró en unificación de doctrina que la relación que une al trabajador y la plataforma digital tenía naturaleza laboral[17]. Posteriormente, en mayo de 2021 se aprueba el Real Decreto-ley 9/2021, de que modifica el Estatuto de los Trabajadores, para garantizar los derechos laborales de las personas dedicadas al reparto en el ámbito de plataformas digitales, más adelante sustituido por la Ley 12 de 28 de septiembre, por la que se modifica el texto refundido de la Ley del Estatuto de los Trabajadores, aprobado por el Real

[16] Bruselas, 21.4.2021, COM(2021) 206 final. 2021/0106 (COD).

[17] STS de 25 de septiembre 2020, casación en unificación de doctrina, ECLI:ES:TS:2020:2924.

Decreto Legislativo 2/2015, de 23 de octubre, para garantizar los derechos laborales de las personas dedicadas al reparto en el ámbito de plataformas digitales.

Esta norma, que trae su origen en un acuerdo de la Mesa del Diálogo Social, ha venido a establecer la presunción de laboralidad de las personas que presten servicios retribuidos consistentes en el reparto o distribución de cualquier producto de consumo o mercancía, por parte de empleadoras que ejercen las facultades empresariales de organización, dirección y control de forma directa, indirecta o implícita, mediante la gestión algorítmica del servicio o de las condiciones de trabajo, a través de una plataforma digital.

Si bien su ámbito de aplicación se circunscribe a la actividad de reparto, y por tanto quedan fuera de la norma cualquier otro tipo de trabajo realizado a través de plataformas digitales, pero el criterio determinante de la presunción de laboralidad sin duda va a ser un sólido referente para la calificación de la relación de trabajo en otros ámbitos que usen igualmente los algoritmos como fórmula de gestión de la relación de trabajo, y para una nueva concepción del concepto de dependencia adaptado a la digitalización del trabajo. Sin restar valor a las iniciativas reguladoras nacionales, hay propuestas doctrinales que sostienen la necesidad de una regulación global del trabajo en plataformas digitales, dado que es un fenómeno global y se extiende ya a escala planetaria, la regulación del mismo debe tener igual escala para ser realmente eficiente (Rodríguez, 2022, 111).

Son otros muchos los aspectos de la relación laboral los que, a través de la IA, el análisis de macrodatos o el internet de las cosas, han sido transformados por la utilización conjunta de estas tecnologías avanzadas y que, en muchos casos, de nuevo chocan con el ejercicio y protección de derechos fundamentales en el trabajo. La utilización de avanzados dispositivos digitales para la vigilancia y control de la ejecución del trabajo, y la geolocalización, tropiezan en no pocos casos con derechos como la intimidad, la protección de datos, la imagen, y ello a pesar de la reciente regulación de esta materia por la Ley Orgánica 3/2018 de protección de datos personales y garantía de los derechos digitales, cuyos resultados normativos son insatisfactorios (Todolí, 2022, 225-226).

Otros aspectos laborales afectados por la digitalización tienen que ver con las condiciones de trabajo, como el tiempo de trabajo, cuya alta dosis de flexibilidad ha ido desdibujando sus contornos, dificultando el concepto de tiempo de trabajo efectivo (retribuido) los tiempos de descanso y la desconexión digital. El tiempo de trabajo es un elemento clave del trabajo, tanto el tradicional, como el informal y el trabajo digital, necesitado de un profundo análisis actual tanto desde el plano jurídico como político y sindical, pues podría ofrecer algunas vías de salida a problemas que a día de hoy están

todavía por resolver, motivados por los nuevos trabajos por tareas mediante plataformas digitales, la sustitución de personas trabajadoras por robots o bots, la digitalización de los procesos de producción y de servicios, o el uso de la inteligencia artificial (Sepúlveda, 2020, 271). En un posible escenario de pérdida masiva de empleo, la reducción del tiempo de trabajo será más necesaria aún que el reparto del tiempo de trabajo (Lozano, 2020, 323). La reducción del tiempo de trabajo es una primigenia reivindicación sindical que hoy día cobra más fuerza que nunca, así como "reclamar entrar en un proceso de negociación tecnológica que permita participar en definir los contornos reales de las condiciones de trabajo futuras del trabajo humano" (Lahera, 2017).

Las nuevas necesidades de formación, la polarización de la ocupación, la brecha digital, los riesgos para la salud, la necesidad de nuevos esquemas para la representación colectiva de los trabajadores, entre otros, son otros tantos efectos laborales de la transformación digital de la economía y que, como los anteriores indicados, exigen una (re)construcción de derechos laborales. La asunción del concepto de equidad digital, como una de las exigencias en las que en el siglo XXI se asienta la dignidad personal y colectiva, constituye una obligación positiva de todos los gobiernos en sus diferentes estructuras estatales, nacionales y locales, y de todas las entidades supranacionales (Cabeza, 2020, 36-37).

4. EL PAPEL DE LA NEGOCIACIÓN COLECTIVA ANTE LA IA Y LA ROBÓTICA

Para Reiner Hoffmann, presidente de la Confederación de Sindicatos Alemanes (DGB) hasta mayo de 2022, la era digital está cambiando rápidamente el mundo del trabajo, y sin embargo son casi exclusivamente las empresas las que se benefician de este "dividendo digital", es decir, de las ganancias de flexibilidad y eficacia que se pueden conseguir con la digitalización. Por ello considera necesario establecer urgentemente barreras de seguridad para el llamado "trabajo inteligente", mediante normas mínimas legales, buenos convenios colectivos y acuerdos de empresa equilibrados (Hoffmann, 2020, 12).

Y es que, en Alemania, la controlabilidad tecnológica ha sido una reivindicación sindical muy temprana, que por medio de la presión colectiva se ha ido consiguiendo paulatinamente en los acuerdos de empresa, hasta que se reconoce mediante ley en los años 70. De este modo, los sindicatos han podido influir desde entonces en el diseño concreto de la tecnología de la empresa respectiva, incluso antes de la instalación de nuevos sistemas tecnológicos -en la industria gráfica- (Uhl, 2020, 6).

Es sabido que la acción sindical no es comparable entre países, por ser algo consustancial a las tradiciones nacionales, la cultura sindical, la idiosincrasia de cada país. Pero este ejemplo nos da idea de la importancia de la negociación colectiva en todo aquello que afecta a los puestos de trabajo y a las condiciones de trabajo, y la digitalización es un determinante clave en ello. El marco legal establece una regulación general, para todos los sectores económicos, pero concierne a la negociación colectiva, como recurso jurídico laboral adaptativo ágil y versátil, dar respuesta idónea a las exigencias de un mercado de trabajo masivamente digitalizado y tecnológico, desde el valor democrático que representa el equilibro de poderes en la gobernanza de las relaciones laborales y la capacidad de ordenar desde su mayor acercamiento a las singularidades de sectores y empresas, aportando flexibilidad y resiliencia, mayor seguridad jurídica y menor nivel de conflictividad (Valdeolivas, 2022, 191).

En esta dirección de participación de los representantes sindicales en la implantación de tecnologías como la IA o la robótica se encamina el Acuerdo Marco Europeo sobre Digitalización, de 22 de junio de 2020, alcanzado por los interlocutores sociales europeos, CES, Business Europe, SME United, y CEEP (y el comité de enlace EUROCADRES/CEC)[18]. Se trata de un pacto ambicioso, que en principio alude a todos los aspectos del mundo del trabajo afectados por la digitalización y sus efectos, y a otros más allá de los estrictamente digitalizados, pues los interlocutores sociales europeos son conscientes de la onda expansiva que produce en todo el ámbito laboral cualquier transformación digital de los procesos de producción de bienes y servicios.

No pretende ser un pacto europeo a modo de código de buenas prácticas para la negociación colectiva en materia de digitalización del trabajo y, sin perjuicio de las diferentes valoraciones que se pueda hacer del mismo, puede suponer un importante impulso para la renovación de contenidos negociales a todos los niveles, y un marco que facilite su adaptación a los cambios en el trabajo como consecuencia de la digitalización de la economía.

El Acuerdo proporciona criterios concretos para la adaptación de contenidos de la negociación colectiva en todos los niveles -empresa, sector, europeo-, motivada por la transformación digital de la economía y, como efecto, del trabajo y de los sujetos que en él intervienen, también les permite delimitar derechos y obligaciones de trabajadores y empresas en algunas concretas materias, a la par que se llama a la intervención negociada sobre las mismas.

[18] Seguimos en este apartado algunos de los resultados recogidos en Sepúlveda Gómez, M. (2021): "El Acuerdo marco europeo sobre digitalización. El necesario protagonismo de la norma pactada", *Revista Temas Laborales* núm. 158/2021.

El Libro Blanco sobre la Inteligencia Artificial, de la Comisión europea, de febrero 2019, COM (2020) 65 final, destaca que los trabajadores y las empresas experimentan las consecuencias directas del diseño y el uso de los sistemas de inteligencia artificial en el lugar de trabajo, por lo que la participación de los interlocutores sociales será un factor decisivo para garantizar un enfoque antropocéntrico de la IA en el trabajo. Este Libro Blanco ha dado lugar a una propuesta de la Comisión europea, presentada al Parlamento europeo el 21 de abril de 2021. Según la comisión especial sobre inteligencia artificial del Parlamento europeo "Europa necesita desarrollar una inteligencia artificial que genere confianza, elimine cualquier tipo de sesgo y discriminación, contribuya al bien común al tiempo que asegure que las empresas y la *industria* prosperan y generan prosperidad económica"[19].

Desde esta perspectiva, el Acuerdo marco se alinea con los objetivos y garantías marcados por las instituciones europeas sobre la IA, y tiene la virtualidad de integrarlos de forma adaptada al ámbito del trabajo y de la negociación colectiva. Entendemos que supone un avance importante, que abre un marco desde la propia negociación colectiva (europea), que marca las directrices a seguir en todos los niveles de negociación en materia de incorporación de la IA al trabajo, y que su actualización puede permitir garantizar los derechos de los trabajadores, en especial, los derechos fundamentales, y aportar fórmulas para el mantenimiento, transición y creación de empleos.

En dicho Acuerdo se destacan las oportunidades y riesgos que para los trabajadores y empleadores puede suponer la IA en el trabajo, así como el aumento de la productividad y del bienestar de los trabajadores, por lo que parte de la premisa de que se deben explorar a través de la negociación colectiva las opciones de diseño de la utilización de la IA, para el éxito económico y las buenas condiciones de trabajo.

El eje central de las directrices del Acuerdo marco descansa sobre el principio del control de las personas sobre la IA en el lugar de trabajo, respetando los controles de seguridad, lo que da lugar a una IA fiable, y se definen los componentes de la misma: Ser lícita, justa, transparente, segura y fiable, cumpliendo con todas las leyes y reglamentos aplicables, así como con los derechos fundamentales y las normas de no discriminación; seguir las normas éticas acordadas, asegurando el respeto de los derechos humanos y fundamentales de la UE, la igualdad y otros principios éticos; ser robusta y

[19] https://www.europarl.europa.eu/news/es/headlines/society/20201015STO89417/regulacion-de-la-inteligencia-artificial-en-la-ue-la-propuesta-del-parlamento.

sostenible, tanto desde el punto de vista técnico como social, ya que, incluso con buenas intenciones, los sistemas de IA pueden causar daños involuntarios.

En consonancia con los condicionantes que debe cumplir la IA fiable, el Acuerdo marco indica alguna de las medidas que se pueden adoptar, que en cierto modo vienen a ser un tanto reiterativas de lo ya dicho con anterioridad, si bien se especifican algunos detalles respecto de alguno de los rasgos de la IA fiable, como son: la necesidad de suministro de información para salvaguardar la transparencia cuando se utilizan sistemas de IA en los procedimientos de recursos humanos, como la contratación, la evaluación, el ascenso y el despido, y el análisis de la actuación profesional. En estos casos se contempla que el trabajador pueda solicitar la intervención humana y/o impugnar la decisión junto con la prueba de los resultados de la IA. O la necesidad de respeto de la legalidad vigente en el diseño y operación de los sistemas de IA, incluido el Reglamento General de Protección de Datos, así como garantizar la privacidad y la dignidad del trabajador.

Cabe destacar, finalmente, que a través de este Acuerdo se configura un modelo de negociación colectiva, dinámico, circular, y permanente, en el que destaca la participación de los representantes de los trabajadores en la toma de decisiones sobre digitalización, implantación, valoración de impactos, y propuestas de revisión. Se pretende concienciar a los negociadores, a todos los niveles, de la importancia de que la digitalización aporte beneficios para todas las personas involucradas en el trabajo, personas trabajadoras y empresas, pero también al empleo y la sostenibilidad del medio ambiente. Una amplitud de fondo y de forma que hace que la negociación colectiva deba afrontar el reto de regenerar sus métodos y contenidos, y una nueva mentalidad que parta de una realidad en proceso de transformación digital y ecológica.

5. CONCLUSIONES

De forma sintética y a modo de cierre, concluimos esta aportación destacando una de las diferencias más palpables entre la primera revolución industrial y la actual revolución digital, desde un punto de vista jurídico-laboral. Y es que, en el ámbito de las relaciones laborales, el punto de partida de la primera revolución industrial fue la de inexistencia de derechos laborales en el trabajo. Estos derechos se han ido construyendo a lo largo de las grandes etapas de cambios, adaptándose a las nuevas realidades sociales y económicas, característica propia del Derecho del Trabajo desde sus orígenes, y han avanzado hasta llegar a un nivel de derechos más o menos aceptable. Pero ese nivel o marco mínimo de derechos se ha ido degradando en una proporción inversa respecto del avance tecnológico, sobre todo digital. La cuarta revolución o industria 4.0 está provocando una inercia de progresivo deterioro de los derechos en el trabajo -sean derechos laborales o derechos fundamentales de la

persona en el trabajo-: las nuevas formas de trabajo, nuevas formas de empresa, nuevos modelos de negocio, nuevas formas de organización de la producción o servicios, están provocando un preocupante incremento de la precariedad laboral, temporalidad, trabajos a demanda, peores condiciones de trabajo, desigualdad, exclusión, desempleo. Lo que resulta paradójico, porque debe ser consustancial al progreso (en este caso tecnológico) comportar mejoras para toda la sociedad.

Desde esta perspectiva, en el actual contexto de digitalización de la economía se habla de la disrupción tecnológica, que representa la idea de que los sistemas progresan creando nuevas estructuras, destruyendo las existentes hasta ese momento. Esta idea o concepto puede suponer un riesgo para el mantenimiento del nivel de derechos alcanzados, como si no fuese ya posible su mantenimiento en el nuevo sistema creado, en el nuevo *statu quo*. En este sentido, es llamativo y muy significativo el uso del lenguaje: se habla de oportunidades y retos de la digitalización de la economía, pero no se incluye en el binomio las oportunidad y riesgos para los derechos en la digitalización. Entendemos que la disrupción de la tecnología no debe equivaler a disrupción de los derechos.

De este modo, el punto de partida para enfrentarnos a estos grandes cambios que supone la digitalización, la IA y la robótica debe ser el mantenimiento del nivel de derechos alcanzados, tanto a nivel internacional como nacional, referidos a las personas, ciudadanos y personas trabajadoras.

Ahora bien, es cierto que estos grandes cambios requieren de una reelaboración de la ordenación jurídica del trabajo, una redefinición de conceptos jurídicos adaptados al nuevo contexto digital, sin que esto suponga un retroceso en el nivel de derechos. Adaptación igualmente de la fuente normativa colectiva, de sus contenidos, de sus objetivos, de forma que la implementación del progreso tecnológico no suponga un retroceso en las condiciones de vida y de trabajo de las personas, sino la co-participación en la construcción del nuevo marco normativo de las relaciones laborales digitalizadas.

6. BIBLIOGRAFÍA

Braña Pino, Francisco Javier (2020), "Cuarta revolución industrial, automatización y digitalización: una visión desde la periferia de la Unión Europea en tiempos de pandemia", en: *Documentos de Trabajo* (Instituto Complutense de Estudios Internacionales) Nueva Época 4.

Cabeza Pereiro, Jaime (2020), "La digitalización como factor de fractura del mercado de trabajo", en: *Revista Temas Laborales* 155, 13-40.

Consejo Económico y Social España (2021): *La digitalización de la economía*. Actualización del informe 3/|2017, consultable en: https://www.ces.es/documents/10180/5250220/Inf0121.pdf/c834e421-ab2d-1147-1ebf-9c86ee56c44a

Cruz García, Paula, Joaquín Maudos Villarroya (2016), "La situación del sector bancario español en el contexto europeo: retos pendientes", en: *Cuadernos Económicos de ICE* 92, 81-108.

Cruz Villalón, Jesús (2017): "Las transformaciones de las relaciones laborales ante la digitalización de la economía", en: *Revista Temas Laborales* 138, 13-47.

Del Rey Guanter, Salvador (dir.), (2018), *Inteligencia artificial y su impacto en los recursos humanos y el marco regulatorio de las relaciones laborales*, Instituto Cuatrecasas de Estrategia Legal en RRHH, Wolters Kluwer, Madrid.

Escande-Varniol, Marie Cécile. (2020), "Relaciones laborales y derechos fundamentales en la era digital", en: *Revista Temas Laborales* 155, 145-189.

Haug, Wolfgang Fritz, (Gustau Muñoz, trad.) (2016), "La digitalización: un cambio de época. El capitalismo de alta tecnología en el umbral de la clausura digital", en: *Pasajes. Revista de pensamiento contemporáneo* 50, 2016, 22-33.

Hoffmann, Reiner (2020), "Mitbestimmung ist für eine humane Gestaltung der Digitalisierung unverzichtbar", *Smart Work!? Mitbestimmung im digitalen Zeitalter* Peter Beule (Hrsg.), Fundación Friedrich-Ebert. https:/www.fes.de/bibliothek/fes-.

Lahera Sánchez, Arturo (2017), "Digitalización, robotización, trabajo y vida: cartografías, debates y prácticas", en: *Cuadernos de Relaciones Laborales* 37, 2, 249-273.

Losano, Mario G. (2006), "La 'iuscibernética' tras cuatro décadas", en: *Cuestiones actuales de Derecho y Tecnologías de la Información y la Comunicación - Revista Aranzadi de Derecho y Nuevas Tecnologías* 4, 15-41.

Lozano Lares, Francisco (2020), "Tiempo de trabajo y derechos digitales", en: Santiago González Ortega (coord.), *El nuevo escenario en materia de tiempo de trabajo*, Temas Laborales - Consejo Andaluz de Relaciones Laborales, Sevilla, 289-326.

Mercader Uguina, Jesús R. (2017), *El futuro del trabajo en la era de la digitalización y la robótica,* Tirant lo Blanch, Valencia.

Ministerio de Industria, Energía y Turismo del Gobierno de España (2015), *La transformación digital de la industria española* (Informe), Estrategia Nacional de Industria 4.0, consultable en: https://www.industriaconectada40.gob.es/estrategias-informes/estrategia-nacional-IC40/Paginas/descripcion-estrategia-IC40.aspx

Ministerio de Trabajo y Economía Social (2021), *Impacto del COVID-19 sobre las estadísticas del ministerio de trabajo y economía social"*, https://www.mites.gob.es/ficheros/ministerio/estadisticas/documentos/Nota_impacto_COVID_DICIEMBRE_2021.pdf.

Purcalla Bonilla, Miguel Ángel (2020), "Seguridad, salud laboral y desconexión digital", en: *Revista Temas Laborales* 155, 109-128.

Rocha Sánchez, Fernando (2017), "La digitalización y el empleo decente en España. Retos y propuestas de adaptación", en: *Futuro del Trabajo: Trabajo decente para todos*, 3, 1-12.

Rodríguez Fernández, María Luz (2022): "Nuevas formas de empleo digital: el trabajo en plataformas. Diez propuestas para su regulación internacional", en: *Digitalización, recuperación y reformas laborales*, Ministerio de Trabajo y Economía Social. Colección Informes y Estudio (Empleo), Madrid.

Sáez Lara, Carmen (2017): "Derechos fundamentales de los trabajadores y poderes de control del empleador a través de las tecnologías de la información y las comunicaciones", en: *Revista Temas Laborales* 138, 185-222.

— (2020), "El algoritmo como protagonista de la relación laboral. Un análisis desde la perspectiva de la prohibición de discriminación", en: *Revista Temas Laborales* 155, 41-60.

Sepúlveda Gómez, María (2019), "Negociación colectiva y derechos digitales en el empleo público", en: *Revista General de Derecho del Trabajo y de la Seguridad Social* 54, 138-167.

— (2020), "El control sindical del tiempo de trabajo", en: González Ortega, Santiago (coord.) *El nuevo escenario en materia de tiempo de trabajo*, Temas Laborales - Consejo Andaluz de Relaciones Laborales, Sevilla, 261-288.

— (2021), "El Acuerdo marco europeo sobre digitalización. El necesario protagonismo de la norma pactada", en: *Revista Temas Laborales* 158, 213-244.

Trillo Párraga, Francisco (2016), "Economía digitalizada y relaciones de trabajo", en: *Revista de Derecho Social* 76, 59-82.

Uhl, Karsten (2020), "Mitbestimmung und Digitalisierung. Die Computerisierung der Druckindustrie in den 1970er-Jahren als Geschichte der Gegenwart", en: Beule, Peter (Hrsg) *Smart Work!? Mitbestimmung im digitalen Zeitalter*, Fundación Friedrich-Ebert. https:/www.fes.de/bibliothek/fes-.

Valdeolivas García, Yolanda (2022), "Derechos de información, transparencia y digitalización", en: *Digitalización, recuperación y reformas laborales*, XXXII Congreso Anual de la Asociación Española de Derecho del Trabajo y de la Seguridad Social, Ministerio de Trabajo y Economía Social, colección Informes y Estudios (Empleo), Madrid.

Valverde Asencio, Antonio José (2019), "Nuevas capacitaciones profesionales para la mejora de la empleabilidad en el proceso de digitalización: Un debate necesario sobre la formación y las políticas activas de empleo", en: *Revista Temas Laborales* 148, 137-160.

— (2020), *Implantación de sistemas de inteligencia artificial y trabajo*, Bomarzo, Albacete.

Widuckel, Werner, Doris Aschenbrenner (2020), "Digital, transformativ, innovativ - Agenda für die Zukunftsfähigkeit Bayerns", *Managerkreis der Friedrich-Ebert-Stiftung Friedrich Ebert Stiftung*, http://library.fes.de/pdf-files/managerkreis/16910.pdf.

World Economic Forum (2022), *Jobs of Tomorrow: The Triple Returns of Social Jobs in the Economic Recovery*, https://www.weforum.org/reports/jobs-of-tomorrow-2022.

CAPÍTULO XVI

INTELIGENCIA ARTIFICIAL Y JUSTICIA DIGITAL[1]

JOSÉ IGNACIO SOLAR CAYÓN
Universidad de Cantabria
jose.solar@unican.es

1. EL DESARROLLO DE LA INTELIGENCIA ARTIFICIAL JURÍDICA Y SU EXPANSIÓN A LA ADMINISTRACIÓN DE JUSTICIA

A estas alturas resulta una obviedad afirmar que la inteligencia artificial se halla presente en prácticamente todos los ámbitos de nuestra actividad cotidiana. Influye en la conformación de nuestros hábitos y preferencias de ocio y consumo (generación automática de ofertas personalizadas, realización de sugerencias cinematográficas y musicales en plataformas digitales, uso de sistemas de reputación para la selección de servicios de alojamiento y restauración…); provoca transformaciones sustanciales en nuestras ocupaciones profesionales (automatización de tareas laborales y rediseño de los procesos de trabajo); determina aspectos importantes para nuestras vidas en áreas como la economía doméstica (la concesión de un crédito o el destino de nuestras inversiones financieras) y la salud (diagnóstico y tratamiento de determinadas enfermedades); y hasta nos auxilia -y, si es preciso, nos suple- en la toma de decisiones vitales (sistemas de asistencia a la conducción, sistemas automáticos de seguridad). Sin duda, estos sistemas "inteligentes", con su enorme potencial de injerencia en nuestra privacidad y su capacidad para modelar nuestras relaciones sociales y esquemas de conducta a través de lo que se ha denominado la "algocracia" (Aneesh, 2009) o "regulación algorítmica" (O'Reilly, 2013), comportan importantes riesgos éticos y jurídicos. Pero también nos facilitan enormemente y, en algunos casos, nos liberan, de multitud de tareas que, de otro modo, nos exigirían mucho tiempo y esfuerzo, siendo además mucho más precisos que cualquier experto humano en su realización.

En el contexto de esta revolución tecnológica, el ciudadano, que hoy tiene al alcance de un *click* en su ordenador o teléfono móvil el acceso inmediato a todo tipo de bienes y servicios en cualquier lugar del mundo, demanda también unos servicios públicos electrónicamente accesibles, ágiles y eficientes. Exigencia que la traumática experiencia social derivada de la pandemia COVID-19 ha revelado una auténtica necesidad. También en el ámbito de la Administración de Justicia, un dominio en el que los procedimientos, los

[1] Este trabajo se ha realizado en el marco del Proyecto de Investigación "La inteligencia artificial jurídica" [RTI2018-096601-B-100 (MCIU/AEI/FEDER, UE)], del Programa Estatal de I+D+i Orientada a los Retos de la Sociedad.

métodos de trabajo y hasta los espacios físicos parecen hasta ahora permanecer impertérritos al paso del tiempo. No resulta difícil vaticinar que, en una sociedad crecientemente digitalizada, el sistema judicial perderá relevancia si no es capaz de adaptarse al cambio tecnológico y de dar una respuesta satisfactoria a las nuevas necesidades, demandas y hábitos sociales. De hecho, creo que esto es algo que ya, en buena medida, está ocurriendo. Cada vez es mayor el número de conflictos jurídicos que encuentran solución al margen del sistema formal de Administración de Justicia, a través de cauces mucho más baratos, rápidos y accesibles. Baste recordar en este sentido cómo la plataforma de adjudicación digital de disputas *Modria* -desarrollada conjuntamente por *eBay* y *PayPal*- resuelve de forma totalmente automatizada, solo entre usuarios de estas compañías, más de 60 millones de disputas cada año: una cifra que supone, aproximadamente, el triple del número total de demandas que recibe en ese mismo período todo el sistema judicial estadounidense (la jurisdicción federal y las cincuenta jurisdicciones estatales)[2].

Además, pese a su conservadurismo secular, lo cierto es que el empleo de sistemas basados en inteligencia artificial es creciente también en el dominio jurídico, y este hecho está comenzando a cambiar sustancialmente la práctica del Derecho. Los extraordinarios avances producidos en los últimos años en las tecnologías de *big data* y en las disciplinas de aprendizaje automático (*machine learning*) y procesamiento del lenguaje natural han permitido el desarrollo de un amplio abanico de sistemas de "inteligencia artificial jurídica", capaces de automatizar o semi-automatizar diversas tareas legales que hasta hace poco era sencillamente inimaginable que pudieran dejar de ser realizadas por profesionales jurídicos expertos.

Este cambio es ya especialmente patente en el ámbito de la abogacía, donde cada vez resulta menos infrecuente que tareas como la investigación legal, la revisión y el análisis de contratos o el seguimiento de su ejecución, la elaboración de dictámenes, la redacción de documentos legales de todo tipo, el asesoramiento legal en determinados ámbitos, la toma de decisiones en relación a la interposición o no de una demanda, la elección de una determinada estrategia procesal o la selección de la información relevante en el litigio, sean realizadas por sistemas basados en inteligencia artificial o por profesionales o para-profesionales asistidos por estos sistemas. Esta automatización o semi-automatización de algunas de las tareas más características de la abogacía

[2] No por casualidad, la compañía *Modria* fue adquirida en 2017 por *Tyler Technologies*, el principal suministrador de tecnología de los tribunales estadounidenses, con el fin de explorar la automatización de diversos tipos de casos. Hoy su plataforma digital de adjudicación de disputas es utilizada en ese país y en Canadá no sólo por asociaciones privadas de arbitraje, sino también por algunas agencias administrativas y tribunales locales.

está teniendo un profundo impacto sobre los métodos de trabajo de estos profesionales y el mercado de servicios jurídicos, propiciando el surgimiento de nuevos modelos de negocio y formas de ejercicio de la profesión, así como una alteración en las formas de acceso y los hábitos de consumo de los servicios legales por parte de los clientes (Solar Cayón, 2019; Barrio Andrés, 2019).

Pero no sólo eso. En tanto algunas de aquellas tareas inciden directamente en el desarrollo del proceso judicial, el empleo de sistemas de inteligencia artificial por parte de la abogacía está comenzando a provocar cambios también en el funcionamiento del sistema judicial e incluso en el rol judicial (Sourdin, 2018, 1115). Un buen ejemplo nos lo proporciona la creciente utilización de los sistemas de análisis predictivo, capaces de detectar, a partir del análisis de los datos extraídos de casos pretéritos, patrones y sesgos en la toma de decisiones por parte de jueces y tribunales. Información que resulta muy valiosa para los profesionales de la abogacía a la hora de determinar cuestiones tales como la probabilidad de éxito de una pretensión, la conveniencia de entablar una demanda (y ante qué tribunal) o de llegar a un acuerdo negociado, o la elección de la estrategia procesal y legal más adecuada en cada caso en función de la composición del tribunal. No se puede desdeñar, pues, la incidencia que tales herramientas puedan tener, a largo plazo, en relación al tipo de disputas que llegan o no a los tribunales (en función de cómo se evalúa el riesgo del cliente) y a la forma en que los asuntos son presentados judicialmente. Pero hay otros posibles efectos más inmediatos. En este sentido, desde algunos gobiernos e instituciones se ha apuntado el riesgo que tales sistemas pueden comportar para la independencia judicial[3]. Y Francia se ha convertido en el primer país que ha limitado legalmente su empleo, sancionando con penas de prisión de hasta cinco años la utilización de los datos de identidad de los magistrados "con el propósito o el efecto de evaluar, analizar, comparar o predecir sus prácticas profesionales reales o presuntas"[4]. Según el *Conseil Constitutionnel*, el objetivo de esta prohibición es evitar presiones sobre los jueces e impedir el mercadeo de estrategias procesales[5].

[3] Sobre estos sistemas, sus potencialidades y riesgos, cfr. Solar Cayón, 2020a, 11-16.

[4] *LOI n° 2019-222 du 23 mars 2019 de programmation 2018-2022 et de réforme pour la justice*, art. 33.

[5] Parece que esta ley constituyó una reacción al proyecto "SupraLegem" del abogado y experto en aprendizaje automático Michaël Benesty, quien desarrolló una plataforma de análisis predictivo para detectar patrones en las sentencias sobre casos de solicitudes de asilo. En 2016 publicó un artículo con los primeros resultados del proyecto, en el que se exponían sesgos bastante llamativos de jueces individualizados -con nombres y apellidos-, mostrando cómo algunos de ellos tenían tasas de rechazo de las solicitudes próximas al 100% (en centenares de casos anuales a lo largo de varios años) mientras que otros pertenecientes al mismo tribunal tenían tasas muy bajas en casos similares. Como era de esperar, el informe suscitó reacciones airadas en la judicatura francesa. Y uno de los primeros efectos que tuvo la ley fue la clausura de la web de "Supralegem".

Otro ejemplo, en este caso con efectos claramente positivos en el proceso judicial, lo encontramos en el empleo de los sistemas de codificación predictiva, desarrollados por la abogacía para seleccionar automáticamente el material electrónico relevante en la fase probatoria del litigio (*discovery*). Tales sistemas son admitidos hoy en la mayoría de las jurisdicciones del *common law*, habiendo agilizado notablemente el desarrollo de esta tarea y provocado importantes reformas procesales y cambios en las funciones del juez, como se expondrá más adelante. De manera que la abogacía está actuando como impulsora de esta transformación tecnológica, alcanzando en su radio de acción también al sistema judicial.

De manera más lenta, la inteligencia artificial está penetrando también en el sector público. En aras de la consecución de una mayor eficiencia en la gestión de los recursos públicos, es creciente el número de gobiernos que utilizan sistemas de inteligencia artificial en diversas áreas de la administración, no sólo para recopilar y analizar datos que ayuden a orientar y formular políticas públicas más eficientes en diversos campos, sino también para tomar de manera automática decisiones que inciden en derechos individuales: detección del fraude fiscal, selección de los beneficiarios de servicios sociales, determinación de los servicios sanitarios a los que tiene derecho cada persona, asignación de becas académicas, evaluación del rendimiento del profesorado y de otros funcionarios, decisiones sobre contratación y renovación de contratos de empleados públicos, y concesión o denegación de visados, entre otras materias[6]. Incluso en algunos ámbitos estamos asistiendo ya a la automatización de procedimientos administrativos sancionadores, como ha sucedido recientemente en nuestro país con la modificación del *Reglamento general sobre procedimientos para la imposición de sanciones por infracciones de orden social y para los expedientes liquidatorios de cuotas de la Seguridad Social* operada por el Real Decreto 688/2021, de 3 de agosto[7].

Y, en la actualidad, buena parte de la atención se está enfocando en la posibilidad de emplear las capacidades de la inteligencia artificial, y de otras tecnologías asociadas como el *big data* y la cadena de bloques (*blockchain*), también en la Administración de Justicia. Son bien conocidos los males que

[6] Sobre los problemas que plantea la utilización de estos sistemas desde el punto de vista de los principios del Derecho Administrativo, cfr. Boix Palop / Cotino Hueso, 2019.

[7] Este Real Decreto incorpora la automatización de la actividad inspectora, permitiendo a la Inspección detectar incumplimientos basados en el análisis masivo de datos sin que se requiera la intervención directa de ningún funcionario. Además, permite el inicio y la tramitación del procedimiento sancionador de forma totalmente automatizada, sin perjuicio de la posibilidad de intervención de personal con funciones inspectoras en la fase de instrucción si fuera preciso hacer una valoración jurídica de las alegaciones efectuadas por los sujetos presuntamente responsables de la infracción.

aquejan el funcionamiento de nuestros sobrecargados tribunales, incluso en los países más avanzados, y la adecuada aplicación de estas tecnologías puede representar una formidable oportunidad para lograr un sistema judicial más ágil, eficiente y accesible a todos los ciudadanos si somos capaces de garantizar el diseño de herramientas respetuosas con las garantías procesales y los derechos del justiciable[8]. Así lo entiende la propia Unión Europea, que ha situado la innovación tecnológica en el centro de su programa de reforma de la Administración de Justicia. En su *2019-2023 Estrategy on e-Justice*, el Consejo de la Unión considera que, particularmente, la inteligencia artificial y el *blockchain* han de constituir áreas de interés prioritario, en cuanto su adecuada utilización podría incrementar la eficiencia y fiabilidad del sistema judicial (Council of the European Union, 2019a, 6). Y, en consonancia con esa estrategia, el *2019-2023 Action Plan European e-Justice* ha definido una serie de proyectos concretos para explorar el papel que estas tecnologías pueden jugar en el diseño de una justicia europea digital[9]. En sintonía con estos planteamientos, en España, la Estrategia "Justicia 2030" contempla como uno de sus pilares esenciales una Ley de Medidas de Eficiencia Digital del Servicio Público de Justicia cuyo actual anteproyecto prevé el uso de la inteligencia artificial y el análisis de *big data* para la automatización de ciertas tareas.

En este contexto, el debate sobre las posibilidades de aplicación de la inteligencia artificial en la Administración de Justicia no ha hecho más que comenzar. Como he expuesto recientemente, cualquier reflexión sobre este asunto precisa, en mi opinión, la adopción de un enfoque pragmático -basado en una apreciación realista de las capacidades y limitaciones de la inteligencia artificial y de sus previsibles efectos beneficiosos y perniciosos-, holístico -que comprenda una visión global del sistema judicial, de sus diversos partícipes y de las distintas funciones y tareas que implica el desarrollo del proceso judicial- y abierto a la complejidad -que tenga en cuenta la heterogeneidad de los problemas suscitados en los distintos contextos procesales por el empleo de los diferentes tipos de sistemas de inteligencia artificial y sus diversas

[8] Esta preocupación motivó la adopción, en diciembre de 2018, de la *European Ethical Charter on the use of Artificial Intelligence in Judicial Systems and their environment* por parte de la Comisión Europea para la Eficiencia de la Justicia (CEPEJ) del Consejo de Europa. La Carta establece cinco principios básicos para el diseño, despliegue y utilización de la inteligencia artificial en el sistema judicial: respeto de los derechos humanos; no discriminación; calidad y seguridad en el procesamiento de las decisiones judiciales y los datos; transparencia, imparcialidad y equidad; y control del usuario. Como veremos a lo largo del trabajo, la Unión Europea también ha ido desarrollando una serie de condiciones y exigencias para el empleo de la inteligencia artificial en sectores de alto riesgo de vulneración de los derechos individuales, como la Administración de Justicia, particularmente en su Propuesta de Reglamento sobre Inteligencia Artificial de marzo de 2021.

[9] Cfr. Council of the European Union, 2019b. En particular, los proyectos nº 11, 12 y 18.

metodologías de diseño y funcionamiento- (Solar Cayón, 2022). En el marco de estas coordenadas, el presente trabajo pretende contribuir a aquel debate exponiendo las principales herramientas de inteligencia artificial que se están empleando en diversas jurisdicciones y las experiencias internacionales más avanzadas en el diseño de tribunales en línea, al objeto de ofrecer al lector una visión panorámica de las posibilidades que estas tecnologías ya ofrecen en su estado actual de desarrollo y de los riesgos que comportan.

2. INTELIGENCIA ARTIFICIAL EN LA AUTOMATIZACIÓN DE TAREAS

Si bien el lector menos informado puede, tal vez, tener la impresión de que hablar de inteligencia artificial en la Administración de Justicia supone adentrarse en el dominio de lo utópico, lo cierto es que una mirada al contexto internacional nos muestra la existencia de un relativamente amplio abanico de sistemas que ya están siendo utilizados para la realización de múltiples y heterogéneas tareas. En algunos casos estas herramientas sustituyen al humano en el desempeño de determinadas funciones o actividades, en otros -la mayoría- le auxilian en su realización y, casi siempre, propician cambios en sus métodos de trabajo, provocando en ocasiones transformaciones importantes en el funcionamiento de los tribunales y en el rol de los participantes en el proceso. A efectos de su sistematización, clasificaré estos sistemas de inteligencia artificial en función de los diferentes tipos de tareas para los que se están empleando, considerados desde la óptica de su incidencia en la función judicial, y particularmente, en la toma de la decisión judicial. Desde este punto de vista podemos distinguir entre tareas meramente instrumentales y auxiliares al proceso, sin incidencia (al menos directa) en la decisión judicial; tareas procesales, en cuanto encaminadas a la realización de diferentes trámites procesales en el transcurso del litigio y/o a la determinación de elementos necesarios para la toma de decisión; y tareas propiamente decisorias de la disputa[10].

[10] La Comisión Europea para la Eficiencia de la Justicia clasifica las aplicaciones tecnológicas en cuatro categorías atendiendo al objetivo que persiguen: acceso a la justicia, comunicación entre los tribunales y los profesionales, administración del tribunal y asistencia directa al trabajo del juez y el secretario (European Commission for the Efficiency of Justice, 2016, 7). Me parece, sin embargo, que la clasificación en función del tipo de tareas que realizan y su incidencia en la resolución judicial resulta más relevante desde un punto de vista jurídico, en cuanto es más afín al enfoque basado en el riesgo (riesgo, fundamentalmente, de vulneración de los derechos de las partes en el proceso) que promueve la Unión Europea en su regulación de la inteligencia artificial. Además, frecuentemente, una misma aplicación sirve simultáneamente diferentes objetivos desde la perspectiva de los distintos implicados en el litigio (juez, profesionales de la oficina judicial, abogados y litigantes, fundamentalmente).

2.1. Inteligencia artificial en tareas auxiliares e instrumentales

Se trata de tareas que facilitan y agilizan el desarrollo del litigio, pero no atañen directamente al ejercicio de las funciones específicamente judiciales. No se trata, siquiera, de tareas jurídicas sino, fundamentalmente, de tareas de carácter administrativo, que tienen que ver con el proceso de digitalización de los expedientes judiciales y su gestión, la tramitación electrónica de los procedimientos y las comunicaciones con los diversos participantes en el proceso, y la asistencia en tareas auxiliares (transcripción de textos, traducción, elaboración de documentos, búsqueda y recuperación de información…). Este es el nivel más básico de automatización en la Administración de Justicia y, generalmente, es a este tipo de tareas a las que suele hacerse referencia cuando se habla de la digitalización de los tribunales. Su automatización no plantea problemas relevantes desde un punto de vista jurídico, en cuanto no implican el ejercicio de funciones propiamente procesales ni, mucho menos, decisorias en relación a ningún aspecto del litigio. Es significativo a este respecto que la propuesta de la Comisión Europea de Reglamento sobre inteligencia artificial, pese a considerar la Administración de Justicia como un dominio especialmente sensible por su incidencia sobre los derechos de los ciudadanos, excluye expresamente de la categoría de "alto riesgo" los sistemas empleados para la realización de "actividades administrativas meramente auxiliares que no afectan a la administración de justicia real en casos individuales" (European Commission, 2021, 28)[11].

La digitalización del expediente judicial y su acceso y gestión electrónicos, así como la digitalización de las sentencias y otros contenidos necesarios para el desarrollo del proceso judicial, la tramitación electrónica de los procedimientos y la gestión automática de determinados procesos administrativos constituyen hoy realidades al alcance de la mano, que no plantean problemas desde un punto de vista jurídico y que pueden redundar de manera inmediata en una mejora muy significativa del funcionamiento de la

[11] No obstante, no se puede descartar que determinadas herramientas aplicadas para la realización de tareas auxiliares puedan tener un efecto -incluso sistémico- sobre el correcto ejercicio de las funciones jurisdiccionales, especialmente si aquellas son diseñadas y gestionadas por agentes externos al poder judicial. Un buen ejemplo nos lo proporciona la polémica suscitada en Polonia a raíz de la puesta en marcha en 2018 del *Random Allocation of Cases System*, un sistema algorítmico utilizado por el Ministerio de Justicia -cuyo titular es, a su vez, el Fiscal General- para distribuir los asuntos entre los jueces. El funcionamiento de esta herramienta ha suscitado enormes recelos por el desigual reparto de los casos y, sobre todo, desde el punto de vista de la independencia judicial, por la adjudicación de varios asuntos en los que se juzgaba a políticos importantes a un mismo juez considerado ideológicamente afín al gobierno. Ante la negativa del Ministerio de Justicia a hacer público el código fuente del sistema, una ONG reclamó judicialmente el acceso al mismo, pero los tribunales han rechazado esa pretensión (Mazur, 2021).

Administración de Justicia, en cuanto a agilidad y eficiencia. Además, la digitalización de documentos y procesos constituye el soporte necesario para el posterior desarrollo de sistemas de *machine learning* basados en el análisis de datos, de manera que esta fase representa un presupuesto imprescindible para la implementación de reformas más ambiciosas en el sistema judicial, como la introducción de herramientas de inteligencia artificial para la realización de tareas específicas más avanzadas en otros niveles, el rediseño de los procesos internos de trabajo y de las formas de administración de justicia, o incluso la creación de tribunales *online*.

Una buena muestra de las tareas que la inteligencia artificial es capaz de realizar para llevar a cabo la digitalización de procesos, expedientes judiciales y otros contenidos nos la proporciona el Tribunal de Justicia de la Unión Europea, que está empleando un amplio conjunto de herramientas desarrolladas por su propio *Innovation Lab*: tecnología de reconocimiento óptico de caracteres para digitalizar textos escritos, aplicaciones para la transcripción automática de archivos de audio y video a texto, y sistemas de reconocimiento automático de voz para generar registros escritos de las sesiones orales del tribunal en tiempo real. También ha desarrollado una plataforma de búsqueda de jurisprudencia que incluye las decisiones digitalizadas de las jurisdicciones nacionales, del propio Tribunal de Justicia de la Unión Europea, del Tribunal Europeo de Derechos Humanos y de la Oficina Europea de Patentes. Además, la gestión de estos contenidos digitalizados requiere el soporte de otras aplicaciones instrumentales, como tecnologías de *big data* para la creación de datos estructurados y su análisis, o herramientas de inteligencia artificial para la clasificación de la información y su posterior búsqueda, selección y recuperación. Al margen de este proceso de digitalización de contenidos, este Tribunal cuenta con sistemas basados en inteligencia artificial para el desarrollo de otras tareas de asistencia tanto a agentes del propio tribunal, como un traductor automático para transcribir los documentos al francés, como a los litigantes, que pueden obtener orientación sobre determinadas cuestiones legales y procesales simples a través de la conversación con un *chatbot*. Dentro de estas tareas auxiliares, otro tipo de herramientas muy interesantes que se están introduciendo en algunos tribunales son las aplicaciones para la generación automática de los diversos tipos de documentos procesales, tanto por parte de los litigantes como del juez y otros agentes de la administración de justicia.

El elemento básico para la digitalización de contenidos y procesos es la implantación de un sistema de gestión de casos (o sistema de gestión procesal, en la terminología española) eficiente para recopilar, gestionar, compartir y enviar la información relativa al proceso entre los distintos partícipes y usuarios del sistema judicial, tanto internos (jueces, personal de la oficina judicial) como

externos (fundamentalmente los diferentes profesionales intervinientes en el litigio -abogados, procuradores, fiscales, expertos…- y las partes), de manera que en cualquier momento cualquiera de ellos pueda conocer el estado del proceso y tenga a su disposición las herramientas necesarias para llevar a cabo a través de una plataforma electrónica la aportación de información o las actuaciones que sean precisas. Este sistema puede incorporar aplicaciones para llevar a cabo múltiples funcionalidades o tareas, tanto en el *back-office* como en el *front-office* de la Administración de Justicia: acceso de los usuarios externos a los datos y documentos del expediente digital, aportación de datos y documentos al proceso por parte de tales usuarios, elaboración automatizada o semiautomatizada de documentos electrónicos a partir de la información digitalizada, racionalización en la gestión de los flujos de información y la organización del trabajo de la oficina judicial, envío de comunicaciones electrónicas a usuarios externos y a instituciones o entidades… Asimismo, en la era del *big data*, es posible incorporar en el sistema de gestión de casos herramientas de minería de datos para monitorizar el funcionamiento de los tribunales y analizar toda la información generada en el curso de su actividad, al objeto de tener un mejor conocimiento de la realidad del sistema judicial y ayudar en la toma de decisiones a la hora de formular las políticas públicas más adecuadas para su mejora (en el aspecto interno), así como de elaborar estadísticas y gráficos para la visualización de datos que procuren (de cara al exterior) una mayor transparencia en el funcionamiento de la Administración de Justicia.

Este es el estadio de la digitalización en el que, a día de hoy, se está actuando en la mayoría de las jurisdicciones europeas, especialmente tras la experiencia de la crisis derivada de la pandemia del COVID-19[12]. Y en el que se encuentra también la Administración de Justicia en España, donde, hasta ahora, el objetivo de la implantación de las nuevas tecnologías se ha limitado a la digitalización de los expedientes y la tramitación electrónica de los procedimientos (Perez Daudí, 2020, 377). Esta situación queda reflejada de manera patente en el *Study on the use of innovative technologies in the justice field* de la Comisión Europea, en el que se recogen todos los proyectos de

[12] La Comisión Europea, en su Comunicación *Digitalisation of justice in the European Union. A toolbox of opportunities*, afirma que "la pandemia del COVID-19 ha subrayado la necesidad de que la UE acelere las reformas nacionales para digitalizar la gestión de casos por parte de las instituciones judiciales, el intercambio de información y documentos con las partes y los abogados, y el acceso fácil y continuo a la justicia para todos", llamando a llevar a cabo dicha tarea "a toda velocidad" (European Commission, 2020a, 1 y 2). Consecuentemente, en el Mecanismo de Recuperación y Resiliencia, aprobado el 11 de febrero de 2021 y que constituye el eje central del plan de recuperación de la Unión Europea *Next Generation EU*, considera la transformación digital de la Administración de Justicia como uno de los sectores en los que se alienta fuertemente a los Estados a enfocar las reformas e inversiones.

inteligencia artificial que están funcionando o se están desarrollando en los sistemas judiciales de la Unión Europea. En relación a España, dicho estudio recoge siete proyectos del Ministerio de Justicia y el Centro de Documentación Judicial (CENDOJ), basados en *machine learning* y procesamiento del lenguaje natural, que se hallan aún en fase de desarrollo o simplemente de planeamiento. Por parte del Ministerio de Justicia se da cuenta de un programa para la transcripción automática de archivos de audio y video a texto (en desarrollo), una aplicación para la clasificación automática de documentos (en desarrollo) y un proyecto de identificación biométrica que utiliza visión digital para facilitar el acceso a la información en el procedimiento penal (planeado). Y por parte del CENDOJ se hallan en fase de desarrollo un proyecto de clasificación automática de sentencias para mejorar su búsqueda y enlazarlas con otros documentos relacionados, otro de creación de datos estructurados que permitan la mejor identificación de una persona, una herramienta de *business intelligence* para mejorar la aplicación de búsqueda de documentos de diverso tipo y hacerla más intuitiva, y un programa para la seudonimización automática de sentencias (European Commission, 2020b, 134-135). En la misma línea, en el plano normativo, el actual anteproyecto de la Ley de Medidas de Eficiencia Digital del Servicio Público de Justicia prevé en su artículo 35 la aplicación de la inteligencia artificial para fines de tramitación electrónica de procedimientos judiciales, búsqueda y análisis de datos y documentos, estadística, anonimización y seudonimización de datos y documentos, uso de datos a través de cuadros de mandos, gestión de documentos, autodocumentación y transformación de documentos, y publicación de información en portales de datos abiertos.

Dentro de las tecnologías para la realización de tareas instrumentales y auxiliares se pueden incluir también aquellas herramientas basadas en inteligencia artificial que asisten a los potenciales usuarios de la Administración de Justicia proporcionándoles, de una forma ajustada a sus necesidades específicas, orientación sobre el derecho sustantivo aplicable a su disputa y sus posibles vías de solución (como los asistentes digitales de voz y los *chatbots*[13]). También, ya en el curso del proceso judicial, aquellas aplicaciones web o de

[13] La aplicación de la inteligencia artificial para el diseño de estos nuevos recursos de información customizada, ajustada a las circunstancias específicas del usuario, ha generado un nuevo paradigma tecnológico: la inteligencia en la interfaz. Aquí el usuario ya no tiene que utilizar motores de búsqueda o navegar por un portal y buscar los canales adecuados de información, sino que simplemente interactúa con un interfaz (hablando o chateando) y la tecnología resuelve los problemas mediante conexiones con diferentes sistemas que pueden responder a sus necesidades, y a partir del aprendizaje. Como señala Corvalán, 2018, 303, la inteligencia en la interfaz puede tener un impacto decisivo para garantizar ciertos derechos de acceso, especialmente cuando se trata de personas vulnerables o con discapacidad.

móvil que ayudan a las partes a realizar determinadas tareas (aplicaciones para la elaboración, firma y presentación de documentos digitales; plataformas para alojar y compartir documentos y pruebas; plataformas colaborativas para la gestión de los expedientes digitales y el acceso a los documentos; aplicaciones de mensajería; etc.) o les permiten participar mediante videoconferencia en determinadas actuaciones procesales. Recientemente, como consecuencia de la pandemia, muchos tribunales han impulsado la introducción de este tipo de aplicaciones que facilitan el desarrollo del litigio sin necesidad de que todos los agentes participantes en un determinado trámite se hallen físicamente presentes en la sala de justicia[14].

En este sentido, es importante que las estrategias de digitalización tengan en cuenta estas capacidades de interacción de los litigantes con el tribunal a través del sistema de gestión de casos. Aspecto que es fundamental para la implementación de una justicia digital y que, al menos hasta ahora, ha sido, en general, bastante descuidado. En la mayoría de ocasiones la digitalización se ha orientado hacia el establecimiento de sistemas de gestión interna de los casos, con más o menos funcionalidades (organización y gestión de los expedientes judiciales, distribución de los casos entre los jueces, estructuración y racionalización de los flujos de trabajo en la oficina judicial, registro de los documentos y trámites del caso, agenda y programación automática de trámites procesales, generación de documentos judiciales, gestión de plazos y envío de avisos, generación de notificaciones automáticas, intercambio de información con otros tribunales y con los registros y bases de datos de otras entidades…). Pero el desarrollo de una justicia digital exige la implementación de otro tipo de aplicaciones tecnológicas, como las señaladas anteriormente, que permitan el acceso y la interacción de los litigantes y sus representantes con la infraestructura digital del tribunal[15]. Y el problema, como apunta Tania Sourdin, es que muy

[14] En Sourdin, 2021, 38-41, el lector puede encontrar una tabla que recoge las innovaciones tecnológicas introducidas por tribunales de distintos países de los cinco continentes para dar respuesta a los problemas de funcionamiento originados por la crisis sanitaria.

[15] En Estados Unidos, el *Institute for the Advancement of the American Legal System* (IAALS), en su informe *Eighteen Ways Courts Should Use Technology to Better Serve Their Costumers*, ha diseñado la arquitectura de la próxima generación de tribunales tecnológicamente avanzados, con el punto de mira puesto en la prestación de un mejor servicio a los usuarios de la Administración de Justicia. El informe establece 18 áreas de desarrollo tecnológico orientadas hacia el exterior para facilitar el acceso de los usuarios a los servicios de la Administración de Justicia, posibilitar su interacción con la oficina judicial y realizar *online* las distintas fases del procedimiento judicial. La mayor parte de las aplicaciones están orientadas a la realización de tareas instrumentales: posibilitar que los usuarios accedan a información, documentos y servicios a través del móvil; que puedan presentar fotos, vídeos y otra información ante el tribunal a través del móvil; comparecer en el tribunal vía telefónica o videoconferencia; programar sus comparecencias a su conveniencia dentro de las posibilidades del calendario judicial; pagar multas, tasas y otras obligaciones financieras *online*; obtener información e impresos a distancia; simplificar el proceso

(…)

pocos sistemas de gestión de casos han sido diseñados teniendo en cuenta este objetivo de interacción con el exterior (Sourdin, 2021, 116). Esto hace que, dadas además las diferentes tecnologías y softwares, frecuentemente, cuando se trata de incorporar a los sistemas ya operativos estas nuevas aplicaciones orientadas al exterior, no puedan ser integradas debido a los problemas de interoperabilidad[16]. Interoperabilidad que es importante lograr no sólo entre las diferentes aplicaciones del tribunal para el desarrollo de las distintas funcionalidades, internas y externas, sino también entre los sistemas de información del tribunal y de todos los operadores que intervienen en los flujos de información virtual (otros tribunales, abogacía, policía, expertos, registros de entes oficiales, etc.)[17]. De este modo, los sistemas de gestión procesal de casos, concebidos inicialmente como sistemas separados y cerrados para la gestión de los expedientes electrónicos, se han constituido hoy en el corazón de la oficina judicial y en el núcleo de un sistema de información mucho más amplio que puede integrar o conectar múltiples funcionalidades basadas en la importación y exportación de datos generados por otras aplicaciones.

---

de cumplimentación de formularios; generar y presentar documentos electrónicamente; generar automáticamente una orden o un mandato a la finalización del correspondiente trámite procesal; disponer de un portal de triaje que proporcione información a los usuarios sobre su problema legal y les oriente hacia los recursos y la vía más adecuada para su solución; creación de un sistema de envío de mensajes automáticos a los usuarios para guiarles en el proceso y avisarles de los distintas trámites y sus requisitos (Greacen, 2018). La Unión Europea también ha publicado un *Toolkit for supporting the implementation of the Guidelines on how to drive change towards Cyberjustice* con recomendaciones específicas para el diseño e implementación de un nuevo modelo de sistema de gestión de casos orientado a los usuarios (European Commission for the Efficiency of Justice, 2019, 28-34).

[16] Para evitar o mitigar estos problemas, el *Institute for the Advancement of the American Legal System*, en el informe mencionado anteriormente, recomienda la implementación de un modelo de sistema de gestión de casos basado en diferentes componentes o módulos para la realización de las distintas funciones que se comuniquen entre sí a través de un interfaz estándar. De este modo, cada tribunal podrá elegir y ensamblar el conjunto de componentes tecnológicos que le aporte el máximo valor en función de sus necesidades específicas. Además, ello facilita la continua actualización tecnológica del sistema sin tener que cambiarlo completamente.

[17] De la importancia de este factor dan cuenta los problemas que actualmente afronta la digitalización de la Administración de Justicia en nuestro país, donde en los últimos años se ha hecho un esfuerzo importante en innovación tecnológica, pero los resultados prácticos son escasos debido a las deficiencias de interoperabilidad entre los diversos sistemas de gestión procesal existentes. Por un lado, contamos con un sistema matriz, "Minerva NOJ", instaurado por la Oficina Judicial Informática, a través del cual se modula el expediente judicial electrónico y que sirve de soporte a la gestión del procedimiento. Al tiempo, existen diversos sistemas interconectados con diferentes aplicaciones procedimentales. Algunas comunidades autónomas disponen de sus propios sistemas y ello genera constantes problemas de interoperabilidad. Por otro lado, orientado hacia el exterior, disponemos del sistema "LexNet Abogacía", vinculado a "Minerva NOJ" y destinado a facilitar los actos de comunicación entre los tribunales y distintos operadores jurídicos, que también ha generado problemas de interoperabilidad (Bueno de Mata, 2020, 13-15).

Tal vez el proyecto de digitalización de la Administración de Justicia más avanzado en este nivel lo encontramos en China, inmersa actualmente en un profundo proceso de reforma y modernización de su justicia que tiene como principal objetivo la implantación de un modelo de *Smart Court*. Este término, acuñado por el Presidente del Tribunal Popular Supremo de China en 2016, no alude a un tribunal determinado, sino al propósito de diseñar -de manera planificada, centralizada y sistemática- una Administración de Justicia basada en la explotación plena del potencial de tecnologías como internet, computación en la nube, *big data*, *blockchain* y diversas ramas de la inteligencia artificial (*deep learning*, procesamiento del lenguaje natural, reconocimiento óptico de caracteres, reconocimiento de voz y reconocimiento de entidades nominales, entre otras)[18]. El primer paso fue la digitalización de contenidos (sentencias, expedientes judiciales, documentos procesales, etc.). Una vez digitalizada esa información, fue posible dar el salto desde la recopilación al análisis de los datos, lo que ha permitido desarrollar un amplio elenco de aplicaciones tecnológicas (internas y hacia el exterior) interconectadas y orientadas a la digitalización de prácticamente toda la actividad procesal (incluida la grabación de los juicios y su retransmisión por *streaming*, que es legalmente obligatoria desde 2018), ofreciendo a los participantes en el litigio (jueces, fiscales, abogados, partes…) y, en general, al público, servicios inteligentes en línea a lo largo de todo el proceso a través de una serie de plataformas digitales multifuncionales (Stern *et al*., 2021, 521-524).

Hoy, en gran parte de los tribunales del país, las partes y los abogados pueden acceder *online* desde sus propios dispositivos electrónicos o desde terminales ubicadas en los tribunales a asistentes virtuales para obtener asesoramiento legal y a herramientas predictivas para conocer las posibilidades de éxito de su pretensión; consultar toda la información sobre el proceso judicial en el que toman parte; recibir notificaciones y actualizaciones automáticas; y realizar en línea operaciones como la presentación de la demanda y de todo tipo de documentos, la elaboración de documentos procesales, la aportación de pruebas y la asistencia y participación en las vistas. El sistema se halla soportado por la conjunción de un amplio abanico de tecnologías: *chatbots*, sistemas de análisis predictivo, herramientas para la generación automática de documentos, plataforma *blockchain* para el

[18] China es hoy el líder global en el desarrollo de la tecnología legal: de las 933 patentes *lawtech* registradas en 2018 en el mundo, más de la mitad lo fueron en ese país (Thompson / Liu, 2019). El lector interesado puede encontrar información pormenorizada sobre la puesta en marcha de la reforma judicial, así como sobre el amplio abanico de tecnologías desarrolladas y las múltiples funciones que realizan, desde la perspectiva de quien era Presidente del Tribunal Supremo de Shanghai cuando este órgano fue designado centro piloto para liderar el proyecto a nivel nacional, en Cui, 2020.

almacenamiento de evidencias, aplicaciones de mensajería instantánea, sistema de reconocimiento facial (que posibilita la identificación para el registro en la plataforma y la realización de los diversos trámites, así como el control de acceso al edificio del tribunal y sus diferentes áreas), reconocimiento de imágenes (para identificar objetos en imágenes, leer textos e identificar palabras…), reconocimiento del habla, video en la nube para el seguimiento de los juicios en línea, computación en la nube…

El desarrollo de este ecosistema tecnológico permitió lanzar en agosto de 2018 un proyecto de "Tribunal Móvil", que hoy se halla desplegado ya en 12 provincias. Se trata de una *app* (*Mobile Micro Court*) que replica un tribunal físico en el ciberespacio trasladando los procedimientos fuera de la sala de justicia, a un teléfono móvil. En esta aplicación se han integrado muchos de los servicios que se pueden encontrar en las plataformas digitales de los tribunales, como el asesoramiento legal, la estimación de las probabilidades de éxito del caso, la presentación de la demanda, la generación y aportación de documentos y pruebas, el acceso a la información sobre el caso y la recepción de comunicaciones. Además, combinando la captura de audio y video remoto, por un lado, y tecnologías de reconocimiento facial y firma electrónica, por otro, posibilita que las partes participen en actividades procesales y comparezcan virtualmente ante el tribunal a través de la plataforma *WeChat*, la principal red social en China (Chen/Li, 2020, 12-13). Este proyecto ha mostrado su eficacia durante la crisis del COVID-19, contribuyendo sustancialmente a fortalecer la resiliencia de la Administración de Justicia: en marzo de 2020 se presentaron 437.000 nuevos casos a través de la aplicación (un 287% más que en el mes anterior), de los cuales, en más de un 72% el proceso de completar la demanda llevó menos de 15 minutos (Shi/Sourdin/Li, 2021, 12).

2.2. Inteligencia artificial en tareas procesales

En este apartado nos referimos a sistemas utilizados para distintos tipos de tareas, realizadas tanto por las partes como por la autoridad judicial, que atañen directamente al desarrollo del proceso y al ejercicio de funciones específicas de la Administración de Justicia, aunque no tengan un carácter resolutorio de la disputa. En este sentido, su automatización puede afectar al rol y las funciones desempeñadas por los diversos participantes en el contexto del proceso. Además, aunque no se trate de tareas propiamente decisorias, en algunos casos preparan o coadyuvan a la resolución del caso, en cuanto sus resultados pueden orientar, y en ocasiones determinar, ciertos elementos o premisas de la decisión judicial. Por ello, la aplicación de la inteligencia artificial en este ámbito puede plantear algunos problemas jurídicos relevantes en relación a la salvaguardia de determinadas garantías procesales y derechos del

justiciable, especialmente cuando nos movemos en la justicia penal[19]. De hecho, como veremos, algunos de los sistemas empleados en este tipo de tareas se encuentran incluidos expresamente entre los considerados de alto riesgo en la propuesta de Reglamento europeo sobre inteligencia artificial y, con toda seguridad, otros que no son mencionados específicamente también son susceptibles de encuadrarse en dicha categoría por tratarse inequívocamente de "sistemas dirigidos a asistir a las autoridades judiciales en la búsqueda e interpretación de los hechos y el Derecho y en la aplicación del Derecho a un conjunto particular de hechos" (European Commission, 2021, 28). Lo cual significa que su diseño ha de cumplir una serie de especificaciones técnicas y que tanto sus proveedores como usuarios están sujetos a rigurosas obligaciones[20].

2.2.1. Sistemas de codificación predictiva para la selección del material relevante en el proceso

Una de las etapas del litigio en la que cada vez es más frecuente el empleo de diversas herramientas basadas en inteligencia artificial y otras tecnologías asociadas es la fase probatoria, donde estas aplicaciones pueden ser utilizadas para realizar tareas tan importantes como la identificación y selección de información relevante en el proceso, la prueba de determinados hechos o la validación, preservación y autenticación de las evidencias electrónicas. Y, de hecho, puede afirmarse que, hasta este momento, el principal "caso de éxito" en el desarrollo de una inteligencia artificial jurídica dirigida al proceso judicial viene representado por la codificación predictiva, una tecnología diseñada específicamente para seleccionar el material documental -entendiendo por documento, en sentido amplio, cualquier información contenida en un soporte electrónico- relevante en el litigio.

---

[19] Sobre la incidencia de los diferentes sistemas de inteligencia artificial jurídica en el derecho al debido proceso y las garantías procesales, cfr. De Asís Pulido, 2021 y San Miguel, 2021.

[20] Según la propuesta de Reglamento, los diseñadores y proveedores de sistemas de alto riesgo deben cumplir una serie de obligaciones en relación al establecimiento de un sistema de gestión de riesgos (art. 9), la calidad y gestión de los datos utilizados para su desarrollo (art. 10), el mantenimiento de una documentación técnica que proporcione información sobre el sistema -características generales, capacidades y limitaciones, algoritmos, datos, entrenamiento, tests y pruebas de validación, sistemas de gestión de los riesgos, etc.- (art. 11) y de registros que aseguren un nivel adecuado de trazabilidad de los resultados (art. 12), la transparencia y provisión de información a los usuarios (art. 13), los mecanismos de supervisión humana del sistema (art. 14) y su precisión, robustez técnica y ciberseguridad (art. 15). En cuanto a los usuarios -en este caso, fundamentalmente, la autoridad judicial-, deben utilizarlos conforme a las instrucciones de uso, asegurarse que los datos introducidos en el sistema -si es que tienen algún control sobre ellos- son relevantes a la vista del objetivo perseguido, monitorizar el funcionamiento del sistema, conservar los registros generados automáticamente y respetar la normativa sobre protección de datos, entre otras obligaciones (art. 29).

Mediante herramientas de aprendizaje automático activo y de procesamiento del lenguaje natural, la codificación predictiva permite realizar la revisión automatizada de enormes volúmenes de datos registrados en cualquier tipo de formato digital (documentos de texto, imágenes y videos, audios, correos electrónicos, bases de datos, calendarios, hojas de cálculo, programas informáticos, comunicaciones a través de internet, etc.), procedentes de múltiples y heterogéneas fuentes (servidores, ordenadores, discos duros, memorias USB, tabletas, teléfonos móviles, correos electrónicos, CDs y DVDs, cintas de *backup*…), e identificar cualquier tipo de información relevante que pueda ser presentada como evidencia ante el tribunal. Estos sistemas utilizan algoritmos de aprendizaje supervisado que "aprenden" los criterios de relevancia jurídica en relación a un caso concreto mediante el análisis de un subconjunto de documentos estadísticamente significativos que han sido previamente codificados de manera manual (es decir, clasificados cada uno de ellos como "relevante" o "no relevante") por un abogado conocedor del asunto. A partir de dichos ejemplos, y tras un proceso iterativo de entrenamiento y ajuste del algoritmo, el sistema genera un modelo predictivo que posteriormente aplica a todos los documentos del conjunto a revisar, clasificándolos como relevantes o no relevantes y priorizándolos mediante la asignación a cada uno de ellos de un grado concreto de probabilidad de relevancia, lo que permite desechar directamente del litigio aquellos que no alcancen un grado mínimo de probabilidad y circunscribir la revisión manual únicamente a aquellos que superen este filtro de relevancia[21].

Se trata de una herramienta que, en el sistema anglosajón, puede ser utilizada por las partes para satisfacer el deber procesal de *discovery*, que obliga a proporcionar al contrincante toda la información que sea relevante para la determinación de los hechos en los que se funda la acción o la defensa, salvo aquella que esté protegida por algún privilegio procesal. La aplicación de la codificación predictiva para automatizar esta tarea procesal fue admitida jurisprudencialmente por primera vez en los Estados Unidos, a raíz de la decisión en *Da Siva Moore* (2012)[22], y hoy es ya una práctica habitual también en los tribunales de Reino Unido, Australia, Irlanda y Canadá, en muchos procesos civiles, mercantiles, administrativos o laborales en los que se hace necesario revisar ingentes volúmenes de información electrónica.

Los efectos de esta tecnología han sido claramente beneficiosos para el sistema judicial, permitiendo resolver un importante problema procesal. Con la ubicuidad de las TICs en todos los ámbitos de nuestra vida personal y

[21] Sobre el desarrollo de esta tecnología y sus diversas metodologías, cfr. Solar Cayón, 2018.

[22] *Da Silva Moore v. Predicis Groupe & MSL Group*, 11-Civ.-1279-ALC-AJP (S.D. New York, 2012).

profesional, y las inmensas capacidades de generación, registro y almacenamiento de datos de los dispositivos electrónicos, la hiperinflación de información electrónica disponible en cualquier disputa de cierta entidad había complicado extraordinariamente el desenvolvimiento de la fase procesal de *discovery*. Por un lado, incrementando notablemente el tiempo y el esfuerzo necesarios para efectuar la revisión manual de la información, así como su coste económico para las partes[23]. Y, por otro, multiplicando las posibilidades de disputa entre los contendientes sobre cuestiones tales como la extensión de la información a revisar o el método adecuado para realizar la búsqueda, a menudo con el único objetivo de dilatar el proceso o, sencillamente, de dificultar la labor del oponente. Hoy, la codificación predictiva ha demostrado que no sólo es un método de revisión de la información electrónica más rápido, eficiente y proporcional en la relación coste-eficacia, sino también bastante más preciso que cualquier otro método alternativo -incluida la revisión manual por parte de equipos de abogados expertos- cuando se manejan grandes volúmenes de datos.

La historia del éxito de la codificación predictiva es interesante también desde la perspectiva jurídica, porque nos muestra cómo la introducción de la inteligencia artificial en el sistema judicial puede originar cambios importantes en los principios procesales e incluso en el rol de los participantes en el litigio, incluida la autoridad judicial. Y es que la aceptación de esta tecnología ha impulsado una sustancial reforma de las normas que rigen esta importante etapa de la fase probatoria. Fundamentalmente, esto se ha traducido, en primer lugar, en un cambio de paradigma procesal basado en la sustitución de una cultura de la confrontación por una cultura de la cooperación entre las partes que permita el diseño de una metodología consensuada y transparente para el entrenamiento imparcial del algoritmo. Y, en segundo lugar, en un redimensionamiento del principio de proporcionalidad, que se convierte -junto con el de relevancia jurídica- en el principio rector del desarrollo de esta fase procesal para asegurar un proceso eficiente de *discovery* en términos de su relación coste-eficacia. Asimismo, estos cambios han incidido en la función judicial, atribuyéndose al juez un papel más activo en la dirección del *discovery* al objeto de impulsar y articular una pronta comunicación y cooperación entre

[23] Para hacernos una idea de las magnitudes, en el caso *Da Silva Moore*, una demanda por discriminación de género presentada por cinco mujeres contra uno de los mayores grupos empresariales de publicidad del mundo, la compañía demandada -que solicitaba utilizar por primera vez esta tecnología- alegaba que debía examinar un total de aproximadamente 3 millones de correos electrónicos. Como dato indicativo, baste señalar que el coste de la revisión manual de los 40.000 primeros documentos priorizados por orden de relevancia, que eran los que inicialmente el demandado se comprometía a revelar a las demandantes en caso de que se le permitiera utilizar esta tecnología, se estimaba ya en 200.000$, esto es, 5$ por documento.

las partes que posibilite el acuerdo sobre los protocolos de codificación predictiva y el intercambio de información[24].

2.2.2. Inteligencia Artificial y *Blockchain* como medios de prueba y como herramientas para la valoración de los medios de prueba

Dentro aún de la fase probatoria, la inteligencia artificial puede constituir también una herramienta importante de apoyo a la función jurisdiccional en la determinación de la premisa fáctica del razonamiento judicial. De particular relevancia es su incidencia en el proceso penal, donde cada vez es más frecuente que sistemas basados en inteligencia artificial aporten información determinante para la investigación y enjuiciamiento de los hechos delictivos, especialmente en lo que se refiere a la determinación de su autoría.

En este ámbito, la inteligencia artificial, asociada a las técnicas de *big data* y minería de datos, puede utilizarse como una herramienta de búsqueda, selección y análisis de la información ya disponible en diferentes tipos de fuentes y bases de datos[25]. En un contexto social y tecnológico en el que la disponibilidad de todo tipo de datos personales crece exponencialmente a partir de la huella digital que voluntaria o involuntariamente vamos generando en nuestra actividad cotidiana, y en el que se está produciendo -incluso en el plano jurídico- una progresiva difuminación de las fronteras entre los datos públicos y los datos privados, las diligencias de investigación tecnológica para el mapeo y análisis de datos a través de sistemas inteligentes parecen destinadas a ocupar una posición preeminente entre los métodos de investigación judicial (Bueno de Mata, 2020, 1 y ss.). En una dirección similar, la omnipresencia de dispositivos basados en inteligencia artificial que registran automáticamente información sobre nuestra ubicación, parámetros biológicos, comunicaciones, actividad, hábitos y otros aspectos personales, hace que ellos mismos puedan constituirse en fuentes de prueba en el contexto de un proceso judicial, pudiendo por tanto ser aportados al mismo como medios de prueba. Por otra parte, también es cada vez más frecuente la utilización de sistemas basados en inteligencia artificial para la valoración de distintos medios de prueba (Nieva Fenoll, 2018, 79 y ss.). Se incluyen aquí herramientas para identificar o establecer correspondencias entre individuos y materiales o rastros digitales, interpretar materiales o rastros digitales ambiguos, reconstruir hechos a partir de ciertas evidencias, etc.

[24] Sobre el impacto de la codificación predictiva en los roles de las partes y del juez, así como las reformas de las normas procesales que ha provocado, cfr. Solar Cayón, 2019, 158-169.

[25] En nuestro país, el actual Anteproyecto de Ley de Enjuiciamiento Criminal contempla expresamente la posibilidad de que el Juez de Garantías autorice "la utilización de sistemas automatizados o inteligentes de tratamiento de datos para cruzar e interrelacionar la información disponible sobre la persona investigada con otros datos obrantes en otras bases de titularidad pública o privada", siempre que concurran determinados requisitos (art. 516).

Además, es bastante habitual que prácticas estándar en la investigación de los hechos, como, por ejemplo, las pruebas de ADN, el análisis de huellas y otros rastros, las pruebas balísticas o los análisis de audios y de imágenes, utilicen también sistemas basados en inteligencia artificial para la determinación de los resultados. Por ello, no resulta en absoluto exagerado afirmar que, en la actualidad, gran parte de los métodos de trabajo de las ciencias forenses se sostienen en algoritmos.

Si bien es evidente que este tipo de pruebas científicas han mejorado sustancialmente las técnicas de investigación, comportan también ciertos riesgos desde la perspectiva jurídica. Y es una realidad innegable que su expansión está generando cambios profundos en el desarrollo de la actividad probatoria que pueden incidir en los valores y garantías procesales. En este sentido, se ha señalado cómo esta irrupción tecnológica está acentuando el creciente énfasis del Derecho Penal en la seguridad y la prevención de riesgos. La aplicación de tecnologías inteligentes para tareas de vigilancia y prevención de posibles delitos, y la posterior utilización de sus resultados en la investigación policial y la prueba procesal, está provocando un desdibujamiento de la línea divisoria entre la función preventiva y el proceso penal: las herramientas tecnológicas que permitieron detectar riesgos se adentran en la investigación policial para convertirse posteriormente en elementos de prueba procesal, incluso favoreciendo las medidas de seguridad postcondena. De manera que la introducción de estas tecnologías no sólo supone la aparición de nuevas fuentes o elementos de prueba e instrumentos de valoración de la prueba, sino también de un nuevo paradigma de *modus operandi* probatorio que se refleja en aspectos como su forma de incorporación al proceso, la aparición de nuevos sujetos intervinientes en la prueba y los efectos de sus resultados científicos (Barona Vilar, 2021, 502 y 600-602).

Un nuevo *modus operandi* que, precisamente, tiende a elevar a categoría de "prueba plena" esos resultados. Y es que cada vez es más frecuente que estas tecnologías resulten determinantes en la fijación de los hechos, resultando sumamente difícil para las defensas desafiar la validez de sus resultados. Como señala Montserrat de Hoyos, en estos casos se corre el riesgo de que la eficiencia tecnológica (real o presunta) se constituya en un criterio autosuficiente sobre la fiabilidad de la prueba, reemplazando totalmente el juicio humano y dejando prácticamente sin efecto la presunción de inocencia (de Hoyos Sancho, 2021, 7-8)[26]. Además, su empleo puede afectar también a la igualdad de armas entre

[26] Por ello, Nuria Borrás afirma que la utilización de esta tecnología para la valoración de pruebas ha de tener en todo caso un carácter auxiliar, pues "la información arrojada por sistemas de inteligencia artificial debe ser tenida en cuenta por el juez únicamente para su propia valoración posterior, sin darles carta de verdad de forma automática" (Borrás Andrés, 2019, 36). El

(...)

las partes si no se garantiza la adecuada transparencia y el acceso a los datos y los algoritmos del sistema, pues de lo contrario resulta prácticamente imposible cuestionar o impugnar sus resultados. Especialmente cuando se les deniega el acceso a la información necesaria para evaluar su funcionamiento, ya sea en aras a proteger el secreto empresarial de los códigos fuente de los sistemas (cuando los algoritmos son desarrollados por una empresa privada, lo que sucede en la mayoría de las ocasiones)[27] o como consecuencia de prácticas procesales poco transparentes[28]. Como han puesto de manifiesto recientes informes en los Estados Unidos y Reino Unido, la proliferación de este tipo de herramientas en el proceso no ha ido, en general, acompañada de las adecuadas garantías de transparencia, y se advierte una falta de estándares explícitos, de buenas prácticas de referencia y de publicidad en el empleo de estos sistemas algorítmicos; lo que puede suponer un menoscabo del derecho al proceso debido, profundizando el ya de por sí significativo desequilibrio de poder que existe en el proceso penal entre el Estado y el acusado. Por ello, recomiendan reformar las reglas procesales del *discovery* para facilitar la supervisión de estos

problema es que para desafiar la validez de los resultados del sistema es necesario poseer unos conocimientos que, desde luego, el juez no posee, por lo que deberá contar con el asesoramiento de expertos.

[27] En Estados Unidos, *People v. Superior Court* (*Chubbs*), nº. B258569, 2015 WL 139069 (Cal. Ct. App. Jan. 9, 2015) fue el primer caso en el que un tribunal de apelación reconoció un *privilege* procesal en beneficio del secreto empresarial en el campo penal. En él se rechazó el acceso del acusado al código fuente del software que evaluó la probabilidad de que su ADN formara parte de una muestra compleja de varios ADNs mezclados extraída de la escena del crimen. E idéntica situación se ha reproducido en *Commonwealth of Pennsylvania v. Michael Robinson*, Court of Common Pleas, Allegheny County, Pennsylvania (Feb. 4, 2016) y *State v. Fair*, nº 10-1-09274-5 SEA (Wash. Super. Ct. King Cty, Jan. 12, 2017). Posición que, hasta ahora, los tribunales estadounidenses han mantenido también en relación a los códigos fuente de otros sistemas algorítmicos utilizados en la justicia penal, como los sistemas de evaluación de riesgos de reincidencia, a los que me referiré posteriormente.

[28] En otros casos se ha impugnado la utilización de este tipo de sistemas por entender que no se ha proporcionado a la defensa información suficiente sobre su empleo, aunque también con muy poco éxito. Así, en *State v. Hickerson*, 228 So. 3d 251 (La. Ct. App. 2018), el fiscal utilizó, sin conocimiento de la defensa, un sistema algorítmico de evaluación de riesgos, entre cuyas funcionalidades se hallaba la elaboración automática de gráficos que mostraban las conexiones entre distintos individuos a partir de la información disponible en su base de datos, para determinar la pertenencia del acusado a una banda responsable de diversos delitos, cuestión que resultó determinante para su condena. Cuando la defensa tuvo conocimiento de ese hecho solicitó la nulidad del juicio, pero el tribunal rechazó la solicitud argumentando que aquel sistema no jugó ningún papel en la resolución del caso. Del mismo modo, en *Lynch v. State*, 260 So. 3d 1166 (Fla. Dist. Ct. App. 2018), se apelaba la condena impuesta a un acusado cuya identificación se realizó mediante un sistema de reconocimiento facial a partir de una foto de móvil. El software, además del condenado, identificó a otros cuatro posibles sospechosos cuyas fotografías no fueron proporcionadas a la defensa. No obstante, el tribunal de apelación confirmó la condena bajo el argumento de que la defensa no pudo demostrar que el resultado del juicio habría sido diferente si hubiera tenido acceso a las fotografías.

sistemas y asegurar que la defensa tiene acceso a una información plena sobre ellos, que debería incluir aspectos como el tipo de herramienta que ha sido empleada, la finalidad para la que fue diseñada, los datos sobre los que opera y otras especificaciones técnicas[29].

En el contexto de la Unión Europea, el apartado 6 del Anexo III de la propuesta de Reglamento sobre Inteligencia Artificial incluye específicamente entre los sistemas de "alto riesgo" varios tipos de herramientas que las autoridades encargadas de aplicar el Derecho puedan emplear en la prueba de los hechos: polígrafos y herramientas utilizadas para detectar el estado emocional de una persona; sistemas utilizados para llevar a cabo análisis criminalísticos en relación con personas físicas que permitan examinar grandes conjuntos de datos complejos, disponibles en diferentes fuentes o formatos, para detectar modelos desconocidos o descubrir relaciones ocultas en los datos; sistemas para la elaboración de perfiles de personas físicas; y sistemas empleados para la evaluación de la fiabilidad de las pruebas durante la investigación o enjuiciamiento de infracciones penales. Por tanto, de aprobarse finalmente esta propuesta legislativa, en todos estos casos deberán observarse las rigurosas exigencias ya descritas en relación a su diseño y funcionamiento.

Lo que parece incuestionable es que, con las debidas garantías, la inteligencia artificial, en conjunción con tecnologías asociadas como la "cadena de bloques" o *blockchain* -que puede constituir una herramienta fundamental para la validación del material probatorio-, está llamada a desempeñar un papel cada vez más importante en la investigación y prueba de los hechos. Un buen ejemplo del potencial que presenta la conjunción de ambas tecnologías nos lo proporciona el proyecto que actualmente están desarrollando la ONG *Global Legal Action Network* y la Universidad de Swansea (Gales) para probar ante los tribunales británicos el empleo de bombas de racimo en la guerra de Yemen, una iniciativa que tiene vocación de extenderse a otros conflictos armados y jurisdicciones internacionales (Hao, 2020). El proyecto consta de dos fases. La primera fue la creación de una base segura de datos que actualmente ya consta de cientos de miles de horas de videos de ataques aéreos procedentes de todo tipo de fuentes abiertas (imágenes de satélites, grabaciones de periodistas, activistas y población civil con móviles y otros dispositivos electrónicos…). Al objeto de garantizar los principios procesales de publicidad e igualdad de armas de cara a su posible utilización ante los tribunales, para la creación de este repositorio se establecieron una serie de requisitos y protocolos muy estrictos de verificación, autenticación y preservación del material en base a la creación de una cadena de custodia mediante *blockchain*. La segunda fase -en la

[29] Cfr. Richardson / Schultz / Southerland, 2019, en relación a los Estados Unidos, y The Law Society of England and Wales, 2019, en relación al Reino Unido.

que se encuentra actualmente el proyecto- consiste en la selección de pruebas mediante inteligencia artificial. A tal efecto se ha desarrollado y entrenado un sistema basado en *machine learning* para detectar aquellas secuencias de video en cuyos fotogramas aparezcan rastros de bombas de racimo. Hasta el momento los resultados son prometedores: más del 90 por ciento del material seleccionado por el algoritmo y sometido a la posterior verificación de expertos contenía, efectivamente, muestras del empleo de ese tipo de bombas (Baena Pedrosa, 2022). Habrá que estar, pues atentos, al posible recorrido judicial de este asunto, así como al posible empleo de este tipo de técnicas y metodologías en otros conflictos bélicos, como la invasión de Ucrania.

Este potencial de la inteligencia artificial y el *blockchain* ya está siendo aprovechado por algunos tribunales para validar y almacenar en plataformas digitales datos y evidencias electrónicas que pudieran ser relevantes en futuros litigios. El ejemplo más prominente son los *Internet Courts* implantados en diversas provincias chinas a partir de agosto de 2017. Se trata de tribunales especializados con competencia para resolver 11 tipos de disputas relacionadas con internet (comercio electrónico, prestación de servicios en línea, préstamos *online*, controversias sobre derechos de propiedad intelectual en relación a los nombres de dominio…). Como se recoge en el *White Paper on the Aplication of Internet Technology in Judicial Practice* publicado en agosto de 2019 por el *Beijing Internet Court*, estos tribunales han utilizado la tecnología de cadena de bloques, en combinación con *big data* y el almacenamiento en la nube, para construir un ecosistema seguro de evidencias electrónicas mediante el desarrollo de una plataforma digital (*Balance Chain System*) que permite a las partes almacenar y validar este tipo de evidencias, asegurando su preservación, así como registrar y confirmar derechos de propiedad intelectual. El sistema garantiza la trazabilidad de las evidencias y su autenticidad (no falsificación). En caso de disputa, esto reduce el coste para las partes y facilita la admisibilidad judicial de las pruebas. En el momento de publicarse el mencionado Libro Blanco (agosto de 2019) sólo en la plataforma del *Beijing Internet Court* se habían descargado y almacenado casi 6,5 millones de evidencias electrónicas, de las cuales más de 1.300 habían sido utilizadas en centenares de litigios sin que se produjera disputa alguna en relación a la autenticidad de la evidencia (Beijing Internet Court, 2019, 17-18). En cuanto a los tribunales ordinarios, a 31 de octubre de 2019, tribunales de 22 provincias chinas estaban conectados ya a una plataforma nacional de evidencias electrónicas basada en *blockchain* (con más de 194 millones de evidencias almacenadas) que se halla interconectada con 27 instituciones, como el *National Time Service Center*, centros de investigación forense, oficinas notariales, plataformas de resolución negociada de disputas, etc. (Supreme People's Court of China, 2019, 75).

2.2.3. Sistemas algorítmicos de evaluación de riesgos de reincidencia criminal

En ocasiones la tarea judicial no consiste en determinar hechos ya acaecidos sino en evaluar la probabilidad de que se produzcan determinados acontecimientos futuros. Así sucede fundamentalmente en relación a la adopción de algunas medidas cautelares, como la prisión provisional, cuya determinación requiere, entre otras consideraciones, una apreciación judicial acerca del riesgo de reincidencia criminal del acusado. Para auxiliar al juez en dicha tarea, en la justicia penal estadounidense se ha generalizado el empleo de los sistemas algorítmicos de evaluación de riesgos. Se trata de herramientas que inicialmente fueron diseñadas para ser utilizadas en el ámbito de la administración penitenciaria, con el objetivo de ayudar a las comisiones competentes a diseñar programas de tratamiento y de rehabilitación que contemplen las necesidades individuales de cada preso (adicciones, necesidades psicológicas y psiquiátricas, orientación laboral, contexto familiar, etc.), así como a tomar las decisiones adecuadas sobre la concesión o no de la libertad condicional y, en su caso, la determinación del grado adecuado de supervisión de dicho régimen. Sin embargo, en los últimos años su empleo se ha extendido también al proceso judicial. En la actualidad son ya habitualmente empleadas en la mayoría de jurisdicciones estatales durante la fase del *pre-trial* para informar las decisiones judiciales relativas a la adopción de medidas cautelares, particularmente la concesión o no de libertad provisional. Y algunas han ido incluso más allá, permitiendo que sus resultados puedan ser tenidos en cuenta en la fase del *trial* como un elemento más para la determinación y graduación de la sentencia condenatoria. De este modo, la aplicación de estas herramientas predictivas ha ido progresivamente permeando todas las áreas y etapas del sistema penal.

Los sistemas de evaluación de riesgos de reincidencia criminal se basan en modelos estadísticos generados automáticamente mediante técnicas de *machine learning* que, a partir del análisis de grandes volúmenes de datos correspondientes a casos pretéritos, son capaces de detectar una serie de correlaciones entre determinados factores personales y sociales y el riesgo de comisión de futuros delitos. Normalmente, estos modelos suelen tomar en cuenta indicadores relativos a las circunstancias personales del acusado (edad y sexo, nivel de estudios, contexto familiar, situación socio-laboral, consumo de drogas…), elementos socio-demográficos (lugar de residencia, contexto socioeconómico, relaciones sociales…) y su historial judicial (detenciones y delitos previos, historial de violencia, precedentes de incomparecencia ante el tribunal…) como factores predictores del riesgo de reincidencia, asignando a cada uno de ellos un peso relativo en función del análisis de los casos pretéritos. De manera que, ante un caso determinado, estas herramientas proporcionan al juez una evaluación del riesgo de reincidencia del acusado, cuantificando la

probabilidad de que vuelva a cometer un delito y determinando, en función de dicha probabilidad, el nivel de riesgo que ha de atribuirse al individuo (bajo, moderado o alto).

El empleo de estos sistemas ha suscitado objeciones importantes por parte de algunos sectores legales y académicos estadounidenses, que han cuestionado su compatibilidad con el derecho del acusado al debido proceso judicial[30]. Como he analizado con mayor profundidad en otro trabajo, al hilo de la polémica sentencia del Tribunal Supremo de Wisconsin en el caso *State v. Loomis* (2016), en el que se cuestionaba la constitucionalidad del empleo de COMPAS -el sistema de evaluación de riesgos de reincidencia más utilizado por los tribunales estadounidenses-, los aspectos potencialmente más problemáticos de estas aplicaciones tienen que ver con tres cuestiones: la falta de transparencia de los algoritmos, que puede menoscabar el derecho de defensa del acusado; el uso de "perfiles" para tomar decisiones que recaen sobre personas individualizadas, que puede afectar a la presunción de inocencia; y la existencia de posibles sesgos, que podría vulnerar el derecho del acusado a un juicio imparcial y el principio de igualdad (Solar Cayón, 2020b).

La falta de transparencia puede obedecer a razones jurídicas, como la ya mencionada protección del secreto empresarial cuando la propiedad de los algoritmos corresponde a compañías privadas -el caso de COMPAS-. En estos casos, las soluciones pueden ser múltiples y relativamente sencillas: arbitrar mecanismos procesales para que la compañía pueda revelar el algoritmo a la defensa o a expertos con las debidas garantías de confidencialidad, exigir a las compañías que vendan estos sistemas a la Administración de Justicia que renuncien al secreto empresarial, o utilizar sistemas de código abierto[31], entre otras. Pero las razones de la falta de transparencia pueden ser también técnicas, como sucede con aquellos sistemas que emplean redes neuronales de aprendizaje profundo (*deep learning*). Como es sabido, los resultados de estos

[30] Aunque es obligado reseñar que cuentan con el beneplácito de las administraciones, la abogacía (a través de la *American Bar Association*) e instituciones tan relevantes como el *American Law Institute*. Los principales argumentos empleados para su defensa son que estos sistemas pueden contribuir a reducir la población reclusa (y sus costes) mediante la identificación de aquellos casos de "riesgo bajo" de reincidencia en los que es posible decretar la libertad provisional, acortar la duración de la condena o incluso sustituir la pena privativa de libertad por otra alternativa; a facilitar la rehabilitación y reinserción de esos individuos de "bajo riesgo" mediante el diseño de programas personalizados; y a racionalizar y uniformizar la evaluación judicial del riesgo de reincidencia, basándola en datos procedentes de miles de casos previos en lugar de intuiciones subjetivas y valoraciones discrecionales del juzgador.

[31] Sistemas como el *Sentence Risk Assessment Instrument*, aplicado en Pennsylvania, y el *Public Safety Assessment (PSA)*, desarrollado por la Fundación Laura y John Arnold y puesto gratuitamente a disposición de cualquier jurisdicción, que se utiliza ya en Arizona, Kentucky, Nueva Jersey y Utah, además de en un amplio número de tribunales locales.

sistemas no son el producto de fórmulas o reglas de decisión predeterminadas, sino de modelos predictivos que son generados automáticamente por el propio sistema a partir del análisis de la interacción entre cientos o miles de indicadores distribuidos en múltiples capas, y que son modificados constantemente a medida que dispone de nuevos datos y adquiere experiencia en la realización de su tarea. Se trata, por tanto, de cajas negras en las que ni siquiera el acceso al código fuente permite desentrañar el peso relativo que ha tenido cada uno de los indicadores analizados en el resultado de la evaluación del riesgo. De manera que, en este caso, la imposibilidad de interpretar o explicar los resultados debería llevarnos a plantear la idoneidad de los sistemas que emplean algoritmos de aprendizaje profundo para realizar esta tarea, por el impacto negativo que pueden tener en el derecho a la defensa del acusado[32].

En relación al empleo de estos sistemas se plantea también el problema del "perfilado", en tanto las evaluaciones automáticas del riesgo de reincidencia se basan en datos de grupos de población que comparten determinadas características, y no en las circunstancias específicas y singulares del individuo sometido a evaluación. Nos hallamos aquí ante un efecto de la "justicia actuarial". Como explica Dominique Robert, una de las características fundamentales de este enfoque es que reconstruye los fenómenos individuales y sociales como factores de riesgo. Por tanto, la unidad de análisis no es ya el individuo o sujeto biográfico sino un perfil de riesgo. A través de técnicas actuariales, "la identidad individual es fragmentada y reconfigurada en una combinación de variables asociadas con diferentes categorías y niveles de riesgo" (Robert, 2005, 11). Un perfil no se ajusta, por tanto, a las características de ningún individuo específico, sino que es una construcción algorítmica resultante de la combinación de diferentes variables identificadas a partir de coincidencias entre algunos de los atributos del individuo sometido a análisis y los perfiles inferidos a partir de un conjunto de atributos extraídos de una masa de individuos (Yeung, 2018, 515). De manera que se atribuye al individuo una peligrosidad sencillamente por el hecho de compartir ciertas características grupales.

Por último, se halla el delicado y complejo problema de los sesgos. El caso de COMPAS es muy ilustrativo en este punto. Un estudio de la ONG *ProPublica* había denunciado la existencia de un sesgo racial en su algoritmo, indicando que mostraba una tendencia a etiquetar erróneamente a los acusados

[32] El Consejo de la Unión Europea se ha referido específicamente a este tipo de sistemas, indicando que "esta falta de transparencia podría socavar la posibilidad de impugnar efectivamente las resoluciones basadas en sus resultados y, por tanto, vulnerar el derecho a un juicio justo y a la tutela judicial efectiva, y limita los ámbitos en los que estos sistemas puedan utilizarse legalmente" (Council of the European Union, 2020, par. 41).

negros como individuos de alto riesgo de reincidencia con más frecuencia que a los detenidos blancos e, inversamente, a etiquetar erróneamente a los detenidos blancos como individuos de bajo riesgo de reincidencia con más frecuencia que a los detenidos negros (Larson *et al.*, 2016). Esta acusación fue rechazada por la compañía propietaria de COMPAS mediante un informe que señalaba que el análisis de *ProPublica* contenía algunos errores técnicos: por un lado, omitía cualquier consideración de la "tasa base" en la interpretación de sus resultados, no teniendo en cuenta las diferentes tasas de reincidencia realmente existentes entre los detenidos de raza negra (51%) y de raza blanca (39%) en la jurisdicción, como mostraban las estadísticas históricas; y, de otro, ignoraba diversas métricas de "imparcialidad" que COMPAS satisface plenamente, como la "paridad predictiva" -sus resultados poseen valores predictivos similares para individuos de distintos grupos raciales- y la "calibración" -el sistema discrimina entre reincidentes y no reincidentes igualmente bien en los diversos grupos- (Dieterich/Mendoza/Brennan, 2016). Más allá de las complejidades y sutilezas estadísticas, lo que la discusión venía a poner de manifiesto, fundamentalmente, es que el algoritmo parecía reflejar de manera fiel -a partir del análisis de los datos de los casos judiciales previos- el sesgo racista del sistema policial y judicial, como demuestra el hecho de que sus tasas de acierto en la predicción de la reincidencia -como reconocía incluso el estudio de *ProPublica*- eran similares en ambos grupos. Dicho de otra manera: si el sistema judicial está racialmente sesgado, las predicciones sólo podrán ser igualmente precisas para ambos grupos si el algoritmo refleja adecuadamente ese sesgo. Es también significativo a este respecto que COMPAS sí aplica directamente modelos predictivos diferentes para hombres y mujeres, en tanto estas tienen tasas históricas de reincidencia sensiblemente inferiores a las de los varones, por lo que un algoritmo neutro o imparcial desde el punto de vista del género conduciría a una sobreestimación sistemática del riesgo de reincidencia de las mujeres y, por tanto, a resultados injustos para ellas (y el tribunal aceptó esta discriminación como justificada). Lo que muestra también cómo la precisión (y, en definitiva, la justicia) puede requerir, en ocasiones, la toma en consideración de criterios técnicos que, a priori, podrían considerarse jurídicamente discriminatorios[33].

Señalar, finalmente, que en este caso el Tribunal Supremo de Wisconsin, pese a reconocer la existencia de los tres problemas señalados, resolvió la constitucionalidad de la utilización judicial de este tipo de sistemas de evaluación del riesgo de reincidencia siempre que se adopten determinadas cautelas, entre las cuales las principales son el carácter meramente orientativo

[33] Para un análisis más pormenorizado de las complejas cuestiones técnicas que plantea la medición de la imparcialidad, cfr. Solar Cayón, 2020b, 156-166.

-no imperativo- de sus resultados y la prohibición de que la decisión judicial se base únicamente en ellos.

Si volvemos la mirada al ámbito europeo, el Anexo III de la propuesta de Reglamento sobre Inteligencia Artificial menciona expresamente este tipo de sistemas entre los considerados de alto riesgo (par. 6 a), por lo que, en principio, no estaría excluida su utilización siempre que cumplan las especificaciones y exigencias requeridas. Exigencias que, en lo relativo a la transparencia (artículo 13 de la propuesta de Reglamento), parecen incompatibles con el empleo de sistemas basados en algoritmos de aprendizaje profundo por su opacidad. En nuestro país, el actual Proyecto de Ley de Enjuiciamiento Criminal también recoge, en su artículo 485, la utilización de instrumentos de valoración del riesgo de violencia o reincidencia criminal, aunque no hace referencia expresa a los sistemas basados en inteligencia artificial. En todo caso, estas herramientas deberán incluir todos los parámetros estadísticos que permitan evaluar tanto su fiabilidad como su capacidad predictiva; habrán de especificar el tamaño de la población con el que han sido construidos, las variables utilizadas como factores de riesgo, los criterios de medición empleados para ponderar dichos factores, asignando puntuaciones, y el tiempo de validez de la predicción; y tendrán que identificarse los estudios de validación realizados. Sería, sin embargo, deseable que, más allá de estos aspectos técnicos, la regulación determinara claramente también, al menos, para qué finalidades pueden ser utilizados sus resultados y qué valor o carácter han de tener estos para el juez.

2.2.4. Sistemas de búsqueda y análisis de la información jurídica

También en la determinación de las premisas normativas del razonamiento judicial juega un papel importante la inteligencia artificial. Hace ya mucho tiempo que el juez ha incorporado entre sus tareas rutinarias el manejo de las herramientas digitales de búsqueda de la información para obtener el material legal, jurisprudencial y doctrinal sobre el que fundar jurídicamente sus decisiones. Estas herramientas vienen incorporando desde hace aproximadamente una década aplicaciones de *machine learning* y procesamiento del lenguaje natural que han hecho mucho más intuitivos y eficientes los sistemas de búsqueda, de manera que el jurista pueda acceder fácilmente a las disposiciones normativas, sentencias y contenidos doctrinales relevantes en relación a una determinada cuestión legal. De hecho, como ya se ha señalado, algunos de los proyectos basados en inteligencia artificial que en estos momentos están desarrollando en nuestro país el Ministerio de Justicia y el CENDOJ tienen precisamente como objetivo mejorar las funcionalidades de clasificación, seudonimización y recuperación de información de estos sistemas.

Pero este área de la investigación jurídica ha experimentado un extraordinario salto cualitativo con el desarrollo de los sistemas de "búsqueda

de respuestas jurídicas" o *legal question answering*, como ROSS o Alexsei, cuyo origen se remonta a la plataforma cognitiva Watson, de IBM. Estos sistemas, basados en algoritmos de aprendizaje profundo supervisado, pueden -tras un arduo proceso de entrenamiento por expertos en la materia jurídica correspondiente- generar automáticamente dictámenes jurídicamente fundamentados en respuesta a cuestiones legales complejas formuladas en lenguaje natural. Ante preguntas, por ejemplo, como "La provisión de alojamiento gratis por parte de un hermano durante un período de más de 20 años ¿constituye "proveer asistencia" conforme a la sección 57(1) de la Ley de Sucesiones (canadiense), al objeto de determinar si se trata de una asistencia de persona a cargo pagadera por un hermano a otro?" y "Asumiendo que el demandante es considerado una persona a cargo de su hermana testadora bajo la Ley de Sucesiones, ¿qué *quantum* por asistencia de persona a cargo puede esperar si ha estado viviendo gratis en la casa de la testadora veinte años?", Alexsei es capaz de elaborar un memorándum que ofrece una conclusión legal y contiene una detallada argumentación jurídica en respuesta a las cuestiones planteadas, con referencia a las normas y los precedentes jurisprudenciales en los que se fundamenta la conclusión (e hipervínculos a esos materiales)[34].

Cada vez más presentes en la abogacía, hasta el momento apenas han sido, sin embargo, empleados en la Administración de Justicia, donde podrían resultar útiles para proporcionar al juez una primera aproximación legal al caso, en la línea de lo sugerido por el Consejo de la Unión Europea, quien llama a explorar en este ámbito las capacidades de la inteligencia artificial para desempeñar tareas como "el análisis, la estructuración y la preparación de información sobre el objeto de los asuntos" o "la evaluación de documentos jurídicos y sentencias" (Council of the European Union, 2020, par. 35).

2.3. Inteligencia artificial en tareas decisorias

2.3.1. Sistemas de negociación automatizada para la resolución de disputas en línea

Como es bien sabido, los sistemas de resolución de disputas en línea fueron originariamente desarrollados por las grandes plataformas de comercio electrónico ante la necesidad de resolver los conflictos a través de un cauce fácil,

[34] Estas preguntas fueron formuladas a Alexsei y a un abogado experto en la materia en una prueba cuyo objetivo era comparar sus respuestas. Estas fueron sustancialmente idénticas, con la única diferencia de que Alexsei citó un precedente relevante más. Y, en cuanto a la eficiencia, el tiempo empleado por el usuario del sistema en introducir las preguntas (en cada caso se deben introducir separadamente la cuestión legal y los hechos del caso) y revisar el dictamen generado automáticamente por Alexsei fue de media hora, mientras que el abogado necesitó 4 horas para su elaboración. Los pormenores de la prueba pueden encontrarse en https://www.wagnersidlofsky.com/ai-legal-research/.

rápido y barato que pudiera ser utilizado por sus usuarios desde cualquier lugar del mundo y en cualquier momento. En ese entorno digital la resolución de disputas en línea se configuraba como la única opción viable para dar respuesta a este tipo de conflictos. Y estos sistemas han ido evolucionando al ritmo de la innovación tecnológica. En una primera fase, se limitaron a replicar y transponer al espacio virtual los mecanismos presenciales de resolución alternativa de disputas, de manera que las tecnologías de la información y comunicación se utilizaban únicamente como herramientas instrumentales que permitían gestionar el conflicto superando las barreras de espacio y tiempo. Posteriormente, se introdujeron algoritmos que, mediante árboles de decisión, aplicaban una serie de reglas predeterminadas a las distintas situaciones posibles para formular propuestas de resolución. Y, en la actualidad, con el desarrollo del *big data* y el *machine learning*, los sistemas son capaces por sí mismos de recolectar, analizar y reutilizar millones de datos para detectar patrones que permiten generar automáticamente propuestas de resolución de las disputas (Katsh/Rabinovich-Einy, 2017, 33-38).

Efectivamente, hoy la combinación de las técnicas de *big data*, minería de datos e inteligencia artificial (a través, sobre todo, de algoritmos de *deep learning*) permite analizar rápida y eficientemente los datos generados en los millones de interacciones y transacciones entre compradores y vendedores, extrayendo patrones de conducta que una inteligencia humana sería incapaz de discernir, y diseñar softwares que automatizan completamente la resolución de estas disputas. Estos softwares codifican sistemas de negociación colaborativa basados en diversas reglas estratégicas para maximizar el interés común de las partes en sus propuestas de solución. Además, a medida que se incrementa el volumen de datos disponibles y de disputas resueltas, el sistema va aprendiendo a ajustar esas propuestas: identificando y asignando valores a los intereses de las partes, identificando potenciales áreas de acuerdo entre ellas a partir de sus respuestas y preferencias, otorgando prioridades a las distintas opciones en función de las elecciones realizadas por las partes en disputas precedentes, etc. De este modo, grandes volúmenes de disputas pueden ser gestionadas y resueltas de manera completamente automática a un coste muy bajo.

En los últimos años, estas herramientas han dado el salto al sector público. En el ámbito administrativo, es creciente el número de agencias administrativas en distintos países que han introducido sistemas de negociación automatizada para resolver disputas de baja intensidad en materia de protección de consumidores y reclamaciones de baja cuantía, pago de

impuestos, relaciones laborales, etc[35]. Y estos mecanismos están comenzando también a ser empleados en determinadas áreas de la justicia civil. Como afirma Maria R. Covelli, antes de emplear la inteligencia artificial para la toma de decisiones judiciales y de plantearse escenarios hipotéticos sobre la introducción de jueces-robots, parecería conveniente impulsar el desarrollo y la utilización de este tipo de sistemas algorítmicos, que no sólo pueden reducir la carga de trabajo de los jueces sino que, además, promueven la autodeterminación de las personas (Covelli, 2019, 129). Determinadas materias de Derecho de familia, como la negociación de los acuerdos de divorcio y las reclamaciones de deudas monetarias de no muy elevada cuantía, así como reclamaciones en materia de consumo, parecen algunos de los ámbitos más propicios para la operatividad de este tipo de sistemas algorítmicos, y en ellas se han centrado fundamentalmente los proyectos pioneros en este campo. En este sentido, a nivel europeo, cabe resaltar el desarrollo del proyecto *Conflict Resolution with Equitative Algorithms* (CREA), impulsado por el Programa Justicia de la Unión Europea, en el que participan diversas universidades europeas. Su finalidad es el diseño de algoritmos para resolver de manera amistosa disputas que implican un reparto de bienes (divorcios, herencias, disolución de sociedades…), optimizando la propuesta de solución a partir de las preferencias expresadas por cada una de las partes sobre los bienes en disputa[36].

En muchos casos este tipo de sistemas son concebidos como mecanismos extrajudiciales, ubicados en plataformas digitales externas a la Administración de Justicia, y frecuentemente surgidos de iniciativas privadas. De hecho, la propia Comisión Europea para la Eficiencia de la Justicia llama la atención sobre el creciente número de plataformas ODR puestas en marcha en muchos países europeos desde el sector privado, bien sea como complemento o en competencia con el sector público (European Commission for the Efficiency of Justice, 2016, 22). Resulta significativa en este sentido la experiencia de la plataforma digital *Rechtwijzer* en Holanda. Implantada en 2014 dentro de la

[35] Sobre la progresiva expansión de estos sistemas a diferentes sectores, tanto privados como públicos, cfr. Katsch/ Rabinovich-Einy, 2017 y Barnett / Treleaven, 2018. Uno de los ámbitos más recientes en los que se están implantando sistemas automáticos de resolución negociada de disputas en línea es el de los *smart contracts* codificados en plataformas de *blockchain*, un dominio para cuya regulación aún no existe un sistema claro y articulado de reglas jurídicas. Sobre este tema, cfr. Schmitz / Rule, 2019 y Rabinovich-Einy / Katsch, 2019. En esta dirección cabe resaltar la reciente constitución de la *Blockchain Arbitration Society* (BAS), una asociación con más de cincuenta empresas afiliadas cuyo objeto es resolver los conflictos privados que puedan surgir entre los usuarios pertenecientes a una red *blockchain*. Considerada la primera jurisdicción virtual del mundo, su Corte Arbitral emitió el 10 de noviembre de 2021 su primer laudo en relación con una transacción de criptomonedas.

[36] Cfr. https://crea-project.eu.

Administración de Justicia como una vía voluntaria y alternativa al proceso judicial, ofrecía una serie de servicios (algunos de ellos automatizados) de información, negociación y mediación en materias de divorcio, reclamaciones de deuda y arrendamientos. Pero el proyecto fue cancelado en julio de 2017 a la vista de su excesivo coste económico y de sus pobres resultados: se estima que apenas en el 1% de los conflictos en materia de divorcio las partes recurrieron a este mecanismo. Sin embargo, hoy funciona con éxito al margen de la Administración de Justicia, habiendo ampliado incluso su campo de acción a disputas en materia de consumo, reclamación de deudas y relaciones laborales[37].

Pero también existen proyectos exitosos de inclusión de plataformas digitales de negociación automatizada en la Administración de Justicia. Esta opción tiene la ventaja de que, si están integradas en el sistema general de información de los servicios judiciales, en caso de que el proceso de negociación no llegue a buen término el tribunal cuenta ya con la información y la documentación sobre el asunto, de manera que el caso queda ya prácticamente listo para el juicio. Las experiencias más innovadoras en este ámbito vienen proporcionadas por algunos de los tribunales digitales implantados en los últimos años, a los que me referiré más extensamente en el último epígrafe de este trabajo. El primero que integró en su seno un mecanismo de resolución negociada de disputas en línea fue el *British Columbia Civil Resolution Tribunal*, un tribunal en línea que desde 2016 resuelve disputas relativas a reclamaciones monetarias (hasta 5.000$), accidentes de tráfico, condominio y en materia de asociaciones y cooperativas. Antes de que un caso pueda llegar a la fase de adjudicación judicial, las partes han de emprender un proceso de negociación en línea a través de una herramienta de negociación automatizada que estructura su interacción mediante el empleo de plantillas y menús desplegables para explorar la posibilidad de alcanzar un acuerdo (y, si éste no es posible a través de esta negociación automatizada, aún dentro de esta misma etapa entra en juego un facilitador humano que asiste a las partes para intentar lograrlo). Y un procedimiento similar está siendo implantado en el Reino Unido dentro de su plan global de reforma de la Administración de Justicia, actualmente en marcha. Una de las piezas centrales de ese plan es la instauración de un tribunal civil digital para la resolución en línea de reclamaciones en diversas materias, en cuyo seno ya están funcionando desde abril de 2018 las plataformas *Divorce Online* y *Online Civil Money Claims* (para reclamaciones de deuda inferiores a 10.000 libras). En ambos servicios, antes de que el caso tenga que ser resuelto por un juez, las partes son remitidas a un proceso de negociación en línea en el que actualmente se combina el empleo de

[37] Cfr. https://rechtwijzer.nl.

aplicaciones de negociación automatizada con la asistencia de un facilitador[38]. En una fase posterior del desarrollo de este tribunal, cuando se hayan recopilado datos suficientes de los casos resueltos cuyo análisis permita diseñar sistemas de resolución automática de disputas, está previsto el empleo de sistemas de inteligencia artificial que generen de forma plenamente automatizada propuestas de solución. Si bien estos proyectos son los más significativos, hoy hay unos 50 tribunales en Estados Unidos y otros en Holanda, Canadá, Reino Unido, India, Brasil y China que han incorporado herramientas para la resolución amistosa de disputas en línea con diferentes grados de automatización en el procedimiento (Martinez, 2020, 1).

2.3.2. Sistemas para la generación automática de (propuestas de) decisiones judiciales

Nos adentramos aquí en el polémico dominio de las decisiones judiciales automatizadas, donde actualmente asistimos a un intenso debate académico sobre la posibilidad de que sistemas basados en inteligencia artificial puedan sustituir al juez[39]. Debate que ha sido avivado incluso por los anuncios de algunos gobiernos europeos, como Estonia (Niiler, 2019) y Francia (Marissal, 2018), sobre la implantación de jueces-robot para resolver determinados tipos de disputas contractuales, aunque tales proyectos no hayan llegado hasta ahora a materializarse. En un trabajo reciente he tenido la oportunidad de extenderme ampliamente sobre este asunto, poniendo de manifiesto los obstáculos técnicos y jurídicos que impedirían -al menos en el contexto europeo- la suplantación del juez humano (Solar Cayón, 2022, 258-267). Baste recordar aquí la posición de la Unión Europea en este punto. Tanto el Consejo (Council of the European Union, 2020, par. 39) como el Parlamento (European Parliament, 2020, par. 71) y, por último, la Comisión Europea en su Propuesta de Reglamento sobre

[38] A fecha de mayo de 2021, *Divorce Online* recibía el 80% de las solicitudes de divorcio en Inglaterra y Gales. El plazo medio de resolución de las solicitudes de mutuo acuerdo es de 10 días, habiéndose reducido en varias semanas respecto al procedimiento "en papel", y el 86% de los usuarios valoran el servicio como "bueno" o "extremadamente bueno". Cfr. https://www.gov.uk/guidance/hmcts-services-online-divorce-and-financial-remedy (consultado el 20 de abril de 2022). En cuanto al *Online Civil Money Claims*, según las estadísticas relativas al año 2021, las partes alcanzaron un acuerdo en el 53% de los casi 8000 procedimientos de mediación realizados. También en este caso parece que la rapidez (los acuerdos se logran, de media, en un plazo de 24 días, frente a los tres meses en el proceso *off-line*) y el ahorro compensan la falta de interacción humana: el 96 por ciento de los usuarios se manifestaron satisfechos o muy satisfechos con el servicio. Cfr. https://www.gov.uk/government/publications/hmcts-service-online-civil-money-claims/hmcts-service-online-civil-money-claims (consultado el 20 de abril de 2022).

[39] Cfr., solo a modo de muestra, Masuhara, 2017; Sourdin, 2018; Luciani, 2019; Anzalone, 2019; Morison / Harkens, 2020; Re / Solow-Niederman, 2019; Scherer, 2019; Martínez Zorrilla, 2019; Battelli, 2020; Rubim, 2020; Belloso Martín, 2021.

inteligencia artificial, subrayan que las decisiones judiciales deben ser siempre tomadas por seres humanos y no pueden delegarse en sistemas de inteligencia artificial, de manera que, en todo caso, el empleo de sistemas de toma automatizada de decisiones debe ir siempre acompañado de una supervisión humana efectiva. Lo que significa que el juez ha de ser capaz de decidir, en cualquier situación particular, no emplear el sistema de inteligencia artificial o, si no, no tomar en consideración, ignorar o revocar su resultado (European Commission, 2021)[40].

En esta dirección, fuera del contexto europeo, encontramos ya algunas experiencias sobre la utilización de la inteligencia artificial para asistir a los jueces en la toma de decisiones mediante el desarrollo de sistemas capaces de generar automáticamente propuestas de sentencia a partir del análisis de la jurisprudencia. De nuevo, los proyectos más avanzados -desde el punto de vista estrictamente tecnológico- los encontramos en China, en el contexto del desarrollo del programa *Smart Court*. A partir del análisis de los datos contenidos en la plataforma *China Judgments Online*, que publica todas las sentencias de todos los tribunales del país (actualmente cuenta ya con más de 100 millones de sentencias en su base de datos), diversos Tribunales Supremos provinciales han diseñado sistemas de este tipo utilizando técnicas de *big data*, *deep learning* y procesamiento del lenguaje natural, con el propósito declarado de asegurar que casos que versan sobre hechos similares sean resueltos de manera similar. El proyecto más emblemático viene constituido por la aplicación "Juez Sabio" (*Rui Fa Guan*), desarrollada por el Tribunal Supremo de Beijing y que puede ser utilizada por cualquier tribunal de esa provincia: es capaz de identificar las cuestiones legales planteadas en un caso, de buscar y recuperar materiales legales relevantes para su resolución y de elaborar una propuesta de decisión basada en sentencias pretéritas. Sistemas similares han sido implantados en Hainan, Shanghai, Guangzhou, y también en algunos *Internet Courts*, como del de Beijing (Shi/Sourdin/Li, 2021, 9-11; Stern, 2021, 540-542). Si bien los jueces son libres de seguir la propuesta de sentencia generada automáticamente, hay algunos tribunales superiores que están empleando dichos sistemas para monitorizar el funcionamiento de los tribunales inferiores y detectar anomalías en las decisiones judiciales a través de

[40] En sintonía, el anteproyecto de Ley de Medidas de Eficiencia Digital del Servicio Público de Justicia prevé en el apartado j) del artículo 35 la posibilidad de aplicar la inteligencia artificial para "la producción de actuaciones judiciales y procesales automatizadas, asistidas y proactivas". Definiendo el artículo 57 la "actuación asistida" como "aquella para la que el sistema de información de la Administración de Justicia genera un borrador total o parcial de documento complejo en base a datos, que puede ser producido por algoritmos, y puede constituir fundamento o apoyo de una resolución judicial o procesal". Borrador que, en todo caso, habrá de ser validado por el juez, quien puede modificarlo libre y completamente.

un sistema de alerta automática que identifica aquellas que difieren significativamente de otras en casos similares (Chen/Li, 2020, 18-19)[41]. Dado el contexto institucional chino, lo que en teoría se presenta como un mecanismo de gestión de riesgos dirigido a fortalecer la integridad judicial, puede representar, sin embargo, como apunta Tania Sourdin, una amenaza para la independencia judicial (Sourdin, 2021, 190-191), que es sin duda uno de los principales peligros que comportan este tipo de sistemas si no se establecen las condiciones y garantías institucionales adecuadas.

Una experiencia mucho más limitada en su alcance, pero muy interesante, viene constituida por el empleo del sistema PROMETEA en el Tribunal Superior de Justicia de la Ciudad Autónoma de Buenos Aires. Con el apoyo de informáticos y expertos en datos, esta herramienta fue diseñada y puesta en funcionamiento en noviembre de 2017 por la Fiscalía General Adjunta de ese tribunal, que es quien recibe los asuntos judiciales en tercera instancia y, una vez estudiados, formula una propuesta de sentencia que eleva al tribunal, el cual finalmente decide el caso. PROMETEA se emplea para generar automáticamente propuestas de decisión en el área contencioso administrativo, en casos de amparo habitacional -en los que están involucradas cuestiones constitucionales relativas al derecho a una vivienda digna, a la salud integral y a los derechos de personas en situación de vulnerabilidad (discapacidad, niños, personas mayores…)- y otros tipos de amparo (cuestiones de empleo público, ejecuciones fiscales, denegaciones de licencias de taxi…). Temáticas que representan las tres cuartas partes de los casos que llegan al tribunal y sobre las que ya existe una jurisprudencia abundante y relativamente estable.

Se trata de un sistema de *machine learning* supervisado que fue entrenado inicialmente con un *data set* de más de 2.400 sentencias del tribunal, las cuales habían sido previamente mapeadas para agruparlas por una serie de temas y subtemas. Es muy importante remarcar el hecho de que su diseño se basa en la construcción de árboles de decisión mediante los que se representan y categorizan de manera sucesiva los diferentes tipos de situaciones y las correspondientes reglas de decisión jurídica formuladas por la propia Fiscalía, de manera que la propuesta de sentencia generada automáticamente por el sistema es el resultado de la aplicación de dichas reglas[42]. Ello hace que el

[41] Un ejemplo nítido de este enfoque nos lo proporciona el Tribunal Popular Intermedio de Taizhou, en la provincia de Zhejiang, que ha implementado un sistema de alerta que consta de 60 indicadores de riesgo y emite etiquetas de alerta en azul, amarillo y rojo según el nivel de riesgo de actuación judicial incorrecta (Supreme People's Court of China, 2019, 82).

[42] Esta fijación por parte de la propia fiscalía de las reglas de decisión del sistema es un aspecto muy importante porque, como afirma Filippo P. Griffi, si el tribunal debe ser imparcial, "la elaboración del algoritmo que decide una causa no puede ser sustraída al aparato jurisdiccional" (Griffi, 2019, 171).

sistema sea completamente trazable y que la propuesta de resolución de cada caso pueda ser interpretada y explicada de una manera clara y sencilla conforme a tales reglas. Una vez generada la propuesta de decisión es revisada por un fiscal, quien puede modificarla o rechazarla antes de elevarla al tribunal: entre las funcionalidades del sistema está la inclusión en el documento generado de hipervínculos para acceder directamente a las normas, sentencias y documentos en los que se basa la propuesta, lo que permite al fiscal verificar fácilmente sus fundamentos jurídicos e incorporar directamente a la propuesta otros argumentos y referencias. Es importante también subrayar en este aspecto la dinamicidad del sistema, cuya "inteligencia" puede ser actualizada y optimizada continuamente por la fiscalía para mejorar su rendimiento, lo que permite adaptarlo inmediatamente a los cambios de jurisprudencia o de criterio del tribunal.

En cuanto a sus resultados, PROMETEA no sólo ha incrementado notablemente la eficiencia de la fiscalía -y, por tanto, del tribunal-, en cuanto ha pasado de tardar un promedio de tres meses en concluir un expediente con su respectiva propuesta de sentencia a hacerlo en cinco días, sino que no ha supuesto una merma en la calidad de su trabajo. Según expone la propia fiscalía, en el año 2018 el porcentaje de concordancia entre las propuestas generadas automáticamente por PROMETEA y el criterio de los fiscales encargados de la revisión alcanzó el 96%. Y, una vez elevadas al tribunal, las propuestas generadas por el sistema fueron confirmadas en el 100% de los casos (en el año 2017, el último antes de la puesta en funcionamiento de este sistema, las propuestas elaboradas manualmente por la fiscalía fueron confirmadas por el tribunal en el 92% de los casos)[43]. A mi juicio, la experiencia de PROMETEA es muy relevante y significativa porque demuestra no sólo la eficiencia del sistema (cuyo coste no superó los 80.000$) sino también -lo que es más importante- el grado de calidad y fiabilidad que pueden alcanzar las decisiones

[43] Para una información más detallada del desarrollo, las funcionalidades y los resultados de PROMETEA, cfr. Corvalán, 2019 y Estévez / Lejarraga / Fillottrani, 2020. Inspirándose en este proyecto, sus creadores han diseñado un sistema de similares características -"PretorIA"- que desde julio de 2020 es utilizado por la Corte Constitucional de Colombia para auxiliar a sus magistrados en la toma de decisiones sobre la admisibilidad de las acciones de tutela de los derechos fundamentales. Hay que tener en cuenta que este tribunal recibió en el año 2019 más de 620.000 sentencias de tribunales inferiores que fueron objeto de acciones de este tipo (casi 1.700 diarias), de manera que el proceso de selección "manual" de aquellas que requieren una tutela urgente supone una carga ingente de trabajo y una demora de tiempo que en muchos casos llega a desvirtuar el objetivo de la acción. "PretorIA" utiliza un algoritmo de aprendizaje automático supervisado entrenado para leer de manera automática todas las sentencias que ingresan y seleccionar las acciones de tutela que requieren un tratamiento prioritario. Con su empleo se ha logrado reducir drásticamente el tiempo de selección de casos urgentes, pasando de 96 días a… ¡2 minutos!

automatizadas en determinadas áreas jurisdiccionales (básicamente, disputas en las que no se discutan los hechos y relativas a sectores jurídicos bien delimitados en los que existan tendencias jurisprudenciales claras y relativamente estables) mediante el diseño de sistemas algorítmicos transparentes cuyos resultados pueden ser explicados y justificados conforme a razones jurídicas, y en el marco de un contexto institucional que garantiza la independencia judicial y la libre toma de decisiones[44].

3. TRIBUNALES *ONLINE* E INTELIGENCIA ARTIFICIAL

Hasta aquí hemos hablado de automatización -o semi-automatización- de tareas, en tanto las experiencias señaladas representan casos de aplicación de la inteligencia artificial para mejorar la eficiencia de los procesos de trabajo actualmente existentes. Y es que, como indica Tania Sourdin, "hasta la fecha, la mayoría de los tribunales han empleado la tecnología para replicar los sistemas y procesos existentes en lugar de enfocarse en una reforma más profunda de las estructuras y procesos judiciales" (Sourdin, 2021, 94). Pero, como se está viendo en otros sectores -por ejemplo, la abogacía- el carácter verdaderamente disruptivo de estas nuevas tecnologías no reside tanto en su capacidad de automatizar tareas cuanto en su potencial de innovación de los procesos para hacer cosas que antes no eran posibles, cambiando radicalmente las formas en las que son prestados los servicios e introduciendo nuevos métodos de trabajo (Susskind, 2012, 6). Esto es, en su potencial para rediseñar los procesos de administración de justicia y el modo en el que los tribunales funcionan. En este sentido, como afirma la Comisión Europea para la Eficiencia de la Justicia, la innovación tecnológica puede representar un medio, y una oportunidad, para implementar ambiciosas reformas judiciales. Cambios que, en todo caso, han de estar ligados a la promoción de los valores esenciales del proceso judicial (European Commission for the Efficiency of Justice, 2019, 6)[45]. Y que no serán solo el resultado de los desarrollos tecnológicos sino también, en buena medida, de las transformaciones sociales, culturales y políticas que acompañarán tales desarrollos (Sourdin, 2021, 272), pues, en última instancia, la legitimidad de las

[44] Sobre las condiciones técnicas, jurídicas, procesales e institucionales que pueden favorecer el empleo de sistemas algorítmicos para la toma de decisiones judiciales automatizadas cfr. Solar Cayón, 2022.

[45] Para ello, como señala Tania Sourdin, es imprescindible garantizar que los jueces estén comprometidos en el rediseño de los procesos judiciales y el desarrollo de los sistemas tecnológicos. No solo para evitar posibles amenazas al principio de separación de poderes, sino también porque ellos son los más capacitados para dirigir el equilibrado proceso que ha de ser llevado a cabo para asegurar que las garantías fundamentales del *rule of law* son preservadas. A los jueces les corresponde un rol esencial como guardianes de la justicia, debiendo jugar un papel preeminente en el desarrollo, supervision y monitorización de los tribunales en línea (Sourdin, 2021, 206).

nuevas formas de administrar justicia dependerá del grado en que la sociedad esté dispuesta a aceptarlas.

En la actualidad, los proyectos más ambiciosos en lo que se refiere a la utilización de las nuevas tecnologías para diseñar formas innovadoras de administrar justicia vienen constituidos por el desarrollo de algunos tribunales *online* para la resolución de disputas en determinadas materias específicas, fundamentalmente de carácter civil. A diferencia de los procesos de digitalización de tareas y trámites procesales que hemos examinado en apartados anteriores, que permiten que determinadas actuaciones se realicen de manera virtual e incluso que algunos de los participantes en el litigio puedan asistir, comparecer y participar en determinados trámites procesales ante el tribunal a través de medios electrónicos (o incluso todos los participantes, de manera que la vista sea completamente virtual), la noción de tribunal en línea alude a la virtualización del propio tribunal, el cual, como explica Richard Susskind, ya no es un lugar sino un servicio. El tribunal como espacio físico es sustituido por una plataforma digital. Y esta desmaterialización del propio órgano abre un nuevo paradigma que comporta cambios sustanciales en las formas de administración de justicia e incluso en nuestra concepción secular acerca de lo que es -y de lo que puede hacer- un tribunal.

En relación al desarrollo del proceso judicial, la virtualización del tribunal posibilita la interacción asíncrona en línea entre los participantes, que ya no precisan concurrir simultáneamente en un espacio -ya sea físico y/o virtual- y en un momento determinado: los documentos, pruebas y argumentos pueden ser presentados y alojados por las partes en la plataforma digital en cualquier momento (las 24 horas del día, los 365 días del año) y la decisión judicial es emitida también a través de aquella. Y este desplazamiento desde un universo judicial síncrono a uno asíncrono, además de generar nuevas formas de actuación que pueden facilitar la accesibilidad al tribunal y mejorar su eficiencia, supone también una revisión de ciertos principios básicos del funcionamiento de nuestro proceso judicial (por ejemplo, la sustitución del principio de oralidad por el de escritura). Por otra parte, esta desmaterialización permite que el tribunal en línea pueda expandir su alcance más allá del ámbito y la función de los tribunales tradicionales, proveyendo a través de una serie de tecnologías integradas en la plataforma digital servicios adicionales a la adjudicación judicial. El tribunal en línea se convierte así en una especie de tribunal "extendido" o "ampliado", capaz de asumir nuevas funciones jurídicas en relación al tratamiento y resolución de los conflictos (Susskind, 2019, 60-61).

De este modo, se promueve el acceso a la justicia en un doble sentido. Por un lado, el tribunal en línea permite un acceso más fácil, rápido y barato, desde cualquier lugar y en cualquier momento, a los servicios de la Administración de

Justicia. Y, por otro, abre paso a una concepción más amplia de justicia que no se limita a la adjudicación judicial de las disputas, sino que comprende el acceso a una serie de mecanismos y herramientas tecnológicas que empoderan jurídicamente a los ciudadanos, como sistemas expertos que proporcionan información y apoyo legal al ciudadano sobre su problema legal y las posibles vías de solución, herramientas de auto-ayuda que contribuyen a la promoción de la "salud jurídica" y de una cultura de prevención de disputas, aplicaciones para la generación automática de documentos legales, sistemas predictivos y herramientas que posibilitan una solución negociada de las disputas, entre otras posibilidades.

De manera que, aunque la noción de tribunales en línea no está directamente orientada hacia la aplicación de la inteligencia artificial jurídica, sí puede contribuir de manera importante a su desarrollo y expansión en cuanto el diseño de aquellos tribunales (plataformas digitales) comporta la construcción de una estructura tecnológica que puede soportar y facilitar la integración de sistemas inteligentes para la realización de distintas tareas y funciones (Sourdin, 2018, 1120). De hecho, como señalaré a continuación, algunos de estos tribunales ya incluyen en la actualidad diversas herramientas de inteligencia artificial y prevén la adición de nuevos sistemas en posteriores fases de su desarrollo.

El proyecto pionero en este ámbito fue la puesta en marcha en 2016 del *British Columbia Civil Resolution Tribunal* en Canadá, competente para la resolución de disputas relacionadas con propiedades en condominio, reclamaciones monetarias por una cuantía inferior a 5.000 dólares, accidentes automovilísticos, y asociaciones y cooperativas[46]. Este tribunal fomenta un enfoque colaborativo entre las partes, promoviendo que éstas utilicen un amplio abanico de herramientas y procedimientos para resolver sus disputas lo antes posible, sin renunciar en última instancia a la adjudicación judicial. Antes de plantear una demanda, el reclamante ha de usar el "Explorador de Soluciones" de la plataforma: expone su situación y, automáticamente, un sistema experto de inteligencia artificial basado en reglas diagnostica su problema y le proporciona información sobre su posición legal y las distintas opciones de solución disponibles, ofreciéndole una serie de herramientas para resolver por sí mismo el problema. Si no es posible la auto-resolución del problema, entonces el reclamante inicia una demanda. Esta demanda se traslada a las partes implicadas para que presenten una respuesta y se inicia un rápido proceso de negociación entre ellas mediante una herramienta de negociación automatizada. Aún en esta fase de resolución amistosa de la

[46] Se accede al tribunal en: https://civilresolutionbc.ca/.

disputa, si la negociación automatizada no tiene resultado, entra en juego un *case manager* o facilitador humano para intentar lograr un acuerdo entre las partes. En caso contrario, la disputa pasa a la fase de la adjudicación judicial: el procedimiento es rápido porque los documentos y las pruebas aportadas por las partes ya están alojados en la plataforma y, de ser necesaria una vista, se realiza a través de Skype. Al finalizar 2020, este tribunal había resuelto 18.335 disputas (de las cuales sólo 3.466 lo fueron mediante decisión judicial), manteniéndose plenamente operativo durante toda la pandemia de la COVID-19[47].

Esta experiencia ha sido una de las principales fuentes de inspiración del ambicioso programa de reforma de la Administración de Justicia británica (Inglaterra y Gales) actualmente en marcha, que, dotada con un presupuesto de más de 1.200 millones de libras, habrá de estar concluida en abril de 2023[48]. El elemento central de la reforma es un tribunal civil en línea (*Online Solutions Court*) que, una vez esté plenamente operativo, resolverá la mayoría de las disputas civiles[49]. Siguiendo el modelo del precedente canadiense, este tribunal en línea está diseñado conforme a una arquitectura de tres capas que refleja

[47] Cfr. https://civilresolutionbc.ca/crt-statistics-snapshot-december-2020 (consultado el 22 de abril de 2022).

[48] Parte del presupuesto proviene de la venta de edificios de tribunales a medida que se implantan servicios de Administración de Justicia en línea. Hasta finales de marzo de 2019 se habían obtenido por este concepto 124 millones de libras: de los 127 tribunales cerrados hasta ese momento, se habían vendido 114, y se preveía la venta de otros 77, estimándose que los ingresos por este capítulo (258 millones) sufragarán el 22% del coste total de la reforma (HM Courts & Tribunals Service, 2019, 7). Según las previsiones, cuando aquella esté completada se espera que la Administración de Justicia británica haya reducido unos 5.000 empleos a tiempo completo, el número de casos resueltos en tribunales "físicos" en 2.400.000 anuales y su presupuesto anual en 265 millones de libras (HM Courts & Tribunals Service, 2018, 4).

[49] Aunque esta digitalización de los servicios judiciales no se limita únicamente al ámbito civil. En el campo penal, en febrero de 2017, al amparo de la *Criminal Justice & Courts Act 2015*, se puso en marcha el *Single Justice Procedure*, que permite que los casos relativos a determinados delitos menores sin víctimas (superación de los límites de velocidad, conducir sin seguro, impago de ciertas tarifas, no tener una licencia legal de televisión…) y que no conllevan pena de prisión puedan desarrollarse íntegramente y resolverse en línea cuando el acusado se declara culpable o no responde a los cargos. Incluso, yendo un paso más allá, el gobierno británico propuso la introducción de un procedimiento automático de condena en línea que permitiría resolver en línea, de manera inmediata y completamente automatizada, aquellos casos sobre algunos de estos delitos en los que el acusado admitiera su culpabilidad y eligiera expresamente este procedimiento: en estos casos, un sistema basado en inteligencia artificial emitiría automáticamente la condena, consistente en el pago de una determinada suma de dinero con arreglo a unos estándares fijados legislativamente (Ministry of Justice UK, 2017). Este proyecto legislativo quedó paralizado con la disolución del Parlamento y la convocatoria de elecciones generales por parte de Theresa May en junio de 2017 y hasta el momento no ha sido retomado.

perfectamente las distintas funciones que asume en relación a la resolución de las disputas legales:

a) La primera capa, de acceso a la plataforma digital, tiene como objetivo la prevención del conflicto mediante la evaluación del mismo y el asesoramiento legal al potencial demandante. Este nivel se halla completamente automatizado. A través de un interfaz simple, el navegador guía al usuario de un modo interactivo mediante una serie de preguntas estructuradas (utilizando para ello un sistema de inteligencia artificial basado en árboles de decisión) a los contenidos relevantes en el caso específico, ayudándole a categorizar jurídicamente sus pretensiones, a entender el Derecho aplicable así como sus derechos y obligaciones, a conocer las opciones y recursos a su disposición, y conectándole con los servicios (judiciales, legales, sociales y asistenciales, etc.) pertinentes en cada caso para la resolución del problema, incluida la generación y presentación de los documentos oportunos ante dichos servicios.

b) El objetivo de la segunda capa es la contención y resolución amistosa de la disputa. En este nivel existen herramientas de negociación automatizada en línea para que las partes puedan llegar a un acuerdo por sí solas, pero, en caso de ser necesario, puede intervenir un "facilitador" a través de internet e incluso telefónicamente. El facilitador, ejerciendo funciones jurisdiccionales bajo la supervisión de un juez, asistirá a las partes para gestionar la disputa y facilitar un acuerdo entre ellas a través del procedimiento más adecuado en cada caso (mediación, conciliación y arbitraje).

c) La tercera capa es la correspondiente a la resolución judicial de la disputa: si esta no ha podido evitarse ni resolverse amistosamente, se somete a un juez que trabaja en línea. Para ello se establece un proceso estructurado, sin necesidad de concurrencia presencial ni síncrona, que se lleva a cabo a través de internet (con el apoyo de medios telefónicos y video-conferencia, si es necesario). Este proceso se basa en un sistema de *continuous online hearing*, en el que las partes pueden presentar documentos, aportar pruebas y realizar sus argumentaciones sobre el caso, con la intervención activa del juez para guiarles en la explicación y comprensión de sus respectivas posiciones, durante un período razonable de tiempo, de manera que los temas en disputa puedan ser explorados y clarificados. Una vez concluido ese plazo, el juez puede emitir su decisión sin necesidad de una vista (aunque, en caso de considerarlo necesario, puede llevarse a cabo de manera no presencial).

Conforme avanza el desarrollo del programa de reforma, se van implementando diversos servicios judiciales en línea sobre diferentes materias. Una de las primeras áreas en las que se ensayó este rediseño de las formas de administrar justicia fue el Derecho de Familia, con la puesta en marcha en junio de 2017 del servicio *Probate Online*, para la gestión de procedimientos

hereditarios en casos no contenciosos, y en abril de 2018 de la plataforma *Divorce Online*, para la resolución de solicitudes de divorcio y de los asuntos económicos asociados, a los que se ha sumado más recientemente la reforma del proceso judicial e implementación del servicio digital en materia de *Family Public Law and Adoption*[50]. Un hito importante en la reforma fue la implementación en marzo de 2018 del servicio *Online Civil Money Claims* para la resolución de reclamaciones monetarias inferiores a 10.000 libras, que, según los datos correspondientes al año 2021, recibe un promedio de 6.526 demandas mensuales y ha conseguido reducir el plazo medio de resolución de las disputas de los más de tres meses que se tardaba conforme al procedimiento judicial tradicional a menos de un mes en la actualidad[51]. También se halla operativo el *Damages Claims Portal*, para determinadas reclamaciones de indemnización por daños, y en la primera mitad de 2022 está proyectada la puesta en marcha del servicio en línea para la resolución de disputas en materia de *possession* (conflictos entre propietarios y arrendatarios, retrasos en el pago de préstamos e hipotecas…).

Si bien en la puesta en marcha de esta primera generación de tribunales en línea británicos la presencia de sistemas basados en inteligencia artificial es aún muy limitada, al menos en lo que se refiere a tareas propiamente jurídicas, como señala Richard Susskind, uno de los principales impulsores de esta reforma, el potencial de estos sistemas es enorme en su próximo desarrollo, donde está previsto que realicen tareas jurídicas y tomen decisiones que en este momento son asumidas por humanos (Susskind, 2019, 274-275). Especialmente en las dos primeras capas de su arquitectura, donde pueden resultar sumamente útiles herramientas como los sistemas de análisis predictivo para ayudar a los usuarios a predecir las probabilidades de éxito de su reclamación y ayudarles así a elegir la vía más apropiada para su resolución (tal como hemos visto que se están empleando ya en algunos tribunales chinos) y los sistemas de negociación que formulan automáticamente propuestas de solución a partir del análisis de los datos generados en casos previos. Sin excluir tampoco el empleo futuro de sistemas automatizados de toma de decisiones judiciales en determinadas áreas (Susskind, 2019, 277-292)[52].

---

[50] El lector interesado puede encontrar información periódicamente actualizada sobre el desarrollo de los distintos servicios en línea en: https://www.gov.uk/government/collections/hmcts-reform-programme-fact-sheets (consultado el 22 de abril de 2022).

[51] Estadísticas disponibles en: https://www.gov.uk/government/publications/hmcts-service-online-civil-money-claims/hmcts-service-online-civil-money-claims (consultado el 22 de abril de 2022).

[52] También el *British Columbia Civil Resolution Tribunal* está explorando la posibilidad de introducir sistemas de inteligencia artificial para la toma de decisiones judiciales automatizadas en determinados asuntos básicos, de pequeña cuantía y que no impliquen significativas consideraciones de política pública, siendo susceptibles de ser resueltos conforme a criterios

(…)

En definitiva, como consecuencia tanto de la automatización o semiautomatización de determinadas tareas como del rediseño de las formas de administración de justicia propiciados por la innovación tecnológica, el proceso de resolución de conflictos en los tribunales está experimentando tres cambios principales: el primero es la transición de un escenario físico a uno semi-virtual o virtual a través de plataformas digitales que posibilitan, según los casos, la combinación de determinadas actividades procesales en línea con otras presenciales o el desarrollo de procesos completamente en línea; el segundo es la sustitución de intervenciones y decisiones humanas por procesos automatizados de negociación o mediación e incluso de toma de decisiones automatizadas; y el tercero es el cambio de modelos de resolución de conflictos que valoraban la confidencialidad a modelos basados en la recolección, análisis y utilización de datos para diseñar y refinar el funcionamiento de los sistemas automatizados de resolución de disputas e incluso prevenir futuros conflictos (Katsch/Rabinovich-Einy, 2017, 162-163). Cambios que, sin duda, modelarán un nuevo tipo de proceso judicial y afectarán profundamente al desarrollo de la función judicial.

4. BIBLIOGRAFÍA

Aneesh, Aneesh (2009), "Global Labor: Algocratic Modes of Organization", en: *Sociological Theory* 27, 4, 347-370.

Anzalone, Angelo (2019), "¿Robotización judicial? Breves reflexiones críticas", en: *Journal of Ethics and Legal Technologies* 1, 1, 95-114.

Baena Pedrosa, Manuel (2022), "Inteligencia artificial en la persecución de crímenes internacionales", en: Solar Cayón, José Ignacio y Sánchez Martínez, Mª Olga (dir.), *El impacto de la inteligencia artificial en la teoría y la práctica jurídica*, La Ley (Wolters Kluwer), Madrid, 317-336.

Barnett, Jeremy y Treleaven, Philip (2018), "Algorithmic Dispute Resolution - The automation of profesional dispute resolution using AI and blockchain technologies", en: *The Computer Journal* 61, 3, 399-408.

Barona Vilar, Silvia (2021), *Algoritmización del Derecho y de la justicia*, Tirant lo Blanch, Valencia.

Barrio Andrés, Moisés (dir.) (2019), *Legal Tech. La transformación digital de la abogacía*, Wolters Kluwer, Madrid.

jurídicos bien establecidos. En estos casos todo el proceso estaría completamente automatizado, salvo cuando fuera precisa la intervención humana para fijar determinados hechos, y la decisión del sistema sería apelable ante un tribunal "humano" (Masuhara, 2017).

Battelli, Ettore (2020), "La decisión robótica: algoritmos, interpretación y justicia predictiva", en: *Revista de Derecho privado* 38, 45-86.

Beijing Internet Court (2019), *White Paper on the application of Internet technology in judicial practice*.

Belloso Martín, Nuria (2021), "Los desafíos iusfilosóficos de los usos de la inteligencia artificial en los sistemas judiciales: a propósito de la decisión judicial robótica vs. decisión judicial humana", en: Belloso Martín, Nuria (dir.), *Sociedad plural y nuevos retos del Derecho*, Aranzadi, Cizur Menor (Navarra), 327-401.

Boix Palop, Andrés; Cotino Hueso, Lorenzo (coord.) (2019), *Revista General de Derecho Administrativo* 50 (número monográfico sobre "Derecho Público, derechos y transparencia ante el uso de algoritmos, inteligencia artificial y big data").

Borrás Andrés, Nuria (2019), "La verdad y la ficción de la inteligencia artificial en el proceso penal", en: Conde Fuentes, Jesús y Serrano Hoyo, Gregorio (dir.), *La justicia digital en España y la Unión Europea*, Editorial Atelier, Barcelona, 31-39.

Bueno de Mata, Federico (2020), "Macrodatos, inteligencia artificial y proceso: luces y sombras", en: *Revista General de Derecho Procesal* 51, 1-31.

Chen, Benjamin Minhao y Li, Zhiyu (2020), "How will technology change the face of Chinese justice?", en: *Columbia Journal of Asian Law* 34, 1, 1-58.

Corvalán, Juan Gustavo (2018), "Inteligencia artificial: retos, desafíos y oportunidades - Prometea: la primera inteligencia artificial de Latinoamérica al servicio de la justicia", en: *Revista de Investigaçoes Constitucionais* 5, 1, 295-316.

— (2019), *PROMETEA. Inteligencia artificial para transformar organizaciones públicas*, Editorial Astrea - Editorial Universidad del Rosario, Buenos Aires - Bogotá.

Council of the European Union (2019a), *2019-2023 Strategy on e-Justice*, 2019/C 96/04.

— (2019b), *2019-2023 Action Plan European e-Justice*, 2019/C 96/05.

— (2020), *Council Conclusions "Access to Justice - seizing the opportunities of digitalization*, 2020/C 342 I/01.

Covelli, Maria Rosaria (2019), "Dall'informatizzazione della giustizia alla decisione robotica? Il giudice del merito", en: Carleo, Alexandra (ed.), *Decisione robotica*, Il Mulino, Bologna, 125-137.

Cui, Yadong (2020), *Artificial intelligence and Judicial Modernization*, Springer - Shanghai People's Publishing House.

De Asís Pulido, Miguel (2021), "Derecho al debido proceso e inteligencia artificial", en: Llano, Fernando, Joaquín Garrido (eds.), *Inteligencia artificial y Derecho. El jurista ante los retos de la era digital*, Aranzadi, Cizur Menor (Navarra), 67-89.

De Hoyos Sancho, Montserrat (2021), "El uso jurisdiccional de los sistemas de inteligencia artificial y la necesidad de su armonización en el contexto de la Unión Europea", en: *Revista General de Derecho Procesal* 55, 1-29.

Dieterich, William; Mendoza, Christina y Brennan, Tim (2016), *COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity*, Northpointe.

Estévez, Elsa; Lejarraga, Sebastián Linares y Fillottrani, Pablo (2020), *PROMETEA. Transformando la Administración de Justicia con herramientas de inteligencia artificial*, Banco Interamericano de Desarrollo, Washington.

European Commission (2020a), *Digitalisation of justice in the European Union. A toolbox of opportunities*, COM(2020) 540 final.

— (2020b), *Study on the use of innovative technologies in the justice field*.

— (2021), *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*, COM(2021) 206 final.

European Commission for the Efficiency of Justice (2016), *Guidelines on how to drive change towards Cyberjustice*.

— (2018), *European Ethical Charter on the use of Artificial Intelligence in Judicial Systems and their environment*, Council of Europe CEPEJ(2018)14.

— (2019), *Toolkit for supporting the implementation of the Guidelines on how to drive change towards Cyberjustice*, Council of Europe CEPEJ(2019)7.

European Parliament (2020), *Resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies*, 2020/2012 (INL).

Greacen, John M. (2018), *Eighteen Ways Courts Should Use Technology to Better Serve Their Costumers*, Institute for the Advancement of the American Legal System (IAALS).

Griffi, Filippo Patroni (2019), "La decisione robotica e il giudice amministrativo", en: Carleo, Alessandra (ed.), *Decisione robotica*, Il Mulino, Bologna, 2019, 165-175.

Hao, Karen (2020), "Human rights activists want to use AI to help prove war crimes in court", en: *MIT Technology Review*, June 25 (https://www.technologyreview.com/2020/06/25/1004466/ai-could-help-human-rights-activists-prove-war-crimes/).

HM Courts & Tribunals Service (2018), *Early Progress in transforming courts and tribunals*, Report by the Comptroller and Auditor General.

— (2019), *Transforming Courts and Tribunals - A progress update*, Report by the Comptroller and Auditor General.

Katsch, Ethan y Rabinovich-Einy, Orna (2017), *Digital Justice: Technology and the Internet of Disputes,* Oxford University Press, New York.

Larson, Jeff *et al.* (2016), *How We Analyzed the COMPAS Recidivism Algorythm*, ProPublica.

Luciani, Massimo (2019), "La decisione giudiziaria robotica", en: Carleo, Alessandra (ed.), *Decisione robotica*, Il Mulino, Bologna, 63-95.

Marissal, Pierric (2018), "Réforme Belloubet. Des logiciels à la place des juges, mirage de la justice predictive", en: *l'Humanité* (https://www.humanite.fr/reforme-belloubet-des-logiciels-la-place-des-juges-mirage-de-la-justice-predictive-654139).

Martinez, Janet K. (2020), "Designing Online Dispute Resolution", en: *Journal of Dispute Resolution* 1, 1-16.

Martínez Zorrilla, David (2019), "El juez artificial: ¿próxima parada?", en: *Oikonomics* 12, 1-12.

Masuhara, David (2017), "Artificial intelligence and adjudication: some perspectives", en: *Amicus Curiae* 11, 2-15.

Mazur, Joanna (2021), "Automated decision-making systems as a challenge for effective legal protection in European Union Law", en: *European Law Review* 46, 2, 194-210.

Ministry of Justice UK (2017), *Transforming our justice system: assisted digital strategy, automatic online conviction and statutory standard penalty, and panel composition in tribunals*, Government response Cm 9391.

Morison, John y Harkens, Adam (2020), "Algorithmic Justice: Dispute Resolution and the Robot Judge?", en: Moscati, Maria; Palmer, Michael y Roberts, Marian (eds.), *Comparative Dispute Resolution*, Edward Elgar Publishing, 339-352.

Nieva Fenoll, Jordi (2018), *Inteligencia artificial y proceso judicial*, Marcial Pons, Madrid.

Niiler, Eric (2019), "Can AI be a fair judge in Court? Stonia thinks so", en: Wired (https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/).

O'Reilly, Tim (2013), "Open Data and Algorithmic Regulation", en: Goldstein, Brett y Dyson, Lauren (eds.), *Beyond Transparency - Open Data and the Future of Civic Innovation*, Code for America Press, San Francisco, 289-300.

Pérez Daudí, Vicente (2020), "La aplicación de las nuevas tecnologías al proceso: ¿realidad o ficción?", en: Fuentes Soriano, Olga (dir.), *Era digital, sociedad y Derecho*, Tirant lo Blanch, Valencia, 373-397.

Rabinovich-Einy, Orna y Katsch, Ethan (2019), "Blockchain and the Inevitability of Disputes: The Role for Online Dispute Resolution", en: *Journal for Dispute Resolution* 2, 47-75.

Re, Richard y Solow-Niederman, Alicia (2019), "Developing artificially intelligent justice", en: *Stanford Technology Law Review* 22, 242-289.

Richardson, Rashida; Schultz, Jason y Southerland, Vincent (2019), *Litigating Algorithms 2019 US Report*, AI Now Institute.

Robert, Dominique (2005), "Actuarial Justice", en: Bosworth, Mary (ed.), *Encyclopedia of Prisons & Correctional Facilities*, Sage, London, 11-13.

Rubim, Pedro (2020), "Paths to digital justice: judicial robots, algorithmic decision-making, and due process" en: *Asian Journal of Law and Society* 7, 3, 1-17.

San Miguel, Cristina, "La aplicación de la inteligencia artificial en el proceso: ¿un nuevo reto para las garantías procesales?, en: *Ius et Scientia* 7, 1, 286-303.

Scherer, Maxi (2019), *Artificial Intelligence and Legal Decision-Making: The Wide Open? Study on the Example of International Arbitration*, Queen Mary University of London, School of Law Legal Studies Research Paper, Nº 318/2019.

Schmitz, Amy J. y Rule, Colin (2019), "Online Dispute Resolution for Smart Contracts", en: *Journal for Dispute Resolution* 2, 103-125.

Shi, Changqing; Sourdin, Tania y Li, Bin (2021), "The Smart Court: A New Pathway to Justice in China?", en: *International Journal for Court Administration* 12, 1, 1-19.

Solar Cayón, José Ignacio (2018), "La codificación predictiva: inteligencia artificial en la averiguación procesal de los hechos relevantes", en: *Anuario de la Facultad de Derecho de la Universidad de Alcalá* 11, 75-105.

— (2019), *La inteligencia artificial jurídica. El impacto de la innovación tecnológica en la práctica del Derecho y el mercado de servicios jurídicos*, Aranzadi, Cizur Menor (Navarra).

— (2020a), "La inteligencia artificial jurídica: nuevas herramientas y perspectivas metodológicas para el jurista", en: *Revus. Journal for Constitutional Theory and Philosophy of Law* 41, 1-27.

— (2020b), "Inteligencia artificial en la justicia penal: los sistemas algorítmicos de evaluación de riesgos", en: Solar Cayón, José Ignacio (ed.), *Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho*, Universidad de Alcalá - Defensor del Pueblo, Alcalá de Henares, 125-172.

— (2022), "¿Jueces-robot? Bases para una reflexión realista sobre la aplicación de la inteligencia artificial en la Administración de Justicia", en: Solar Cayón, José Ignacio y Sánchez Martínez, Mª Olga (dir.), *El impacto de la inteligencia artificial en la teoría y la práctica jurídica*, La Ley (Wolters Kluwer), Madrid, 245-280.

Sourdin, Tania (2018), "Judge v. Robot? Artificial Intelligence and judicial decision-making", en: *UNSW Law Journal* 41, 4, 1114-1133.

— (2021), *Judges, Technology and Artificial intelligence*, Edward Elgar Publishing.

Stern, Rachel *et al.* (2021), "Automating Fairness? AI in the Chinese Courts", en: *Columbia Journal of Transnational Law* 59, 515-553.

Supreme People's Court of China (2019), *Chinese Courts and Internet Judiciary* (https://www.chinajusticeobserver.com/law/x/chinese-courts-and-internet-judiciary).

Susskind, Richard (2012), *Provocations and Perspectives*, Working paper submitted to the UK CLE Research Consortium (Legal Education and Training Review).

— (2019), *Online Courts and the future of Justice*, Oxford University Press.

The Law Society of England and Wales (2019), *Algorithms in the Criminal Justice System*, The Law Society Commission on the Use of Algorithms in the Justice System.

Thompson, Barney y Liu, Nian (2019), "China leads the way in legal technology patents, new figures show", en: *Financial Times*, February 17.

Yeung, Karen (2018), "Algorithmic Regulation: A Critical Interrogation", en: *Regulation & Governance* 12, 505-523.

IV. Derecho internacional, Estado de derecho y Administración digital

CAPÍTULO XVII

SIGNIFICADO Y ALCANCE DE LOS VALORES DE LA CARTA DE NACIONES UNIDAS EN LA REGULACIÓN INTERNACIONAL DE LA INTELIGENCIA ARTIFICIAL (IA)[1]

DANIEL GARCÍA SAN JOSÉ[2]
Universidad de Sevilla
dagarcia@us.es

1. INTRODUCCIÓN

Este estudio parte de dos premisas (García San José, 2021a; 2021b): la primera premisa es que si bien la inteligencia artificial (en adelante "IA") tiene potencial para transformar el futuro de la humanidad para mejor y en favor del desarrollo sostenible, también existe una conciencia generalizada de los riesgos y desafíos que conlleva, especialmente por lo que respecta a la agravación de las desigualdades y brechas existentes, así como las implicaciones para los derechos humanos (UNESCO, 2019b: punto 11); ahora bien, no existe un marco ético internacionalmente aceptado en relación con todas las aplicaciones e innovaciones de la inteligencia artificial (UNESCO, 2019b: punto 3).

Este hecho puede explicarse, en parte, tras constatar que si bien a nivel doctrinal existe una cierta coincidencia a la hora de definir la IA (Boden, 2016), no ocurre lo mismo en cuanto a la definición de la IA cara a su regulación internacional pudiendo apreciarse diversidad de aproximaciones a su conceptualización a nivel normativo internacional y regional: sistemas de IA (Comisión Europea, 2020; UNESCO, 2021), IA (OCDE, 2019) (Naciones Unidas, 2019), sistemas algorítmicos (Consejo de Europa). De ahí que desde hace años se trabaje en los distintos foros internacionales y regionales en presentar diversos instrumentos reguladores de la IA desde principios éticos asumibles por el mayor número posible de Estados.

En este sentido, pueden citarse como eslabones más relevantes de estos esfuerzos a nivel internacional, dentro de la Unión Europea su *Libro Blanco de la Comisión Europea sobre la Inteligencia Artificial: un enfoque europea orientado hacia la excelencia y la confianza*, COM (2020) 65 final, de 19 de febrero de 2020, que sigue

[1] Estudio realizado en el marco del Proyecto de I+D del Ministerio de Ciencia e Innovación de España *Biomedicina, Inteligencia Artificial, Robótica y Derecho: los Retos del Jurista en la Era Digital*. (PID2019-108155RB-I00), en el que participo como investigador a tiempo completo.

[2] ORCID: 0000-0003-1288-9655. Scopus: 37008234700. Dialnet: 170520. Para comentarios: dagarcia@us.es

las *Directrices para una Inteligencia Artificial fiable* del Grupo de Expertos de Alto Nivel, creado por la Comisión Europea y publicadas un año antes. En la otra gran Organización europea, el Consejo de Europa, se adoptó la *Recomendación CM/Rec (2020) 1 del Comité de Ministros a los Estados Miembros sobre los impactos de los sistemas algorítmicos en los derechos humanos*, el 8 de abril de 2020. A nivel universal, la UNESCO ha aprobado el 23 de noviembre de 2021 su Recomendación sobre la Ética de la IA que comentamos con más detalle en las páginas siguientes de este trabajo. De igual manera, deben citarse los *Principios Éticos de la OCDE en materia de Inteligencia Artificial* adoptados en mayo de 2019, los cuales fueron posteriormente incluidos como anexo a la Declaración Ministerial sobre Comercio y Economía Digital, del G20 en su cumbre de Tsukuba, Japón, en junio de ese mismo año. La OCDE también ha creado un grupo de expertos (AIGO) para que formulen directrices respecto a la especificación de los principios relativos a la inteligencia artificial en la sociedad. En el ámbito de Naciones Unidas, ha de mencionarse la estrategia del Secretario General de esta Organización en materia de nuevas tecnologías, recogida en la Resolución de la Asamblea General A/RES/73/17, de 3 de diciembre de 2018, así como los esfuerzos liderados por la Unión Internacional de Telecomunicaciones en la órbita del sistema de Naciones Unidas (sus principales organismos y agencias), que recoge anualmente en una publicación titulada *United Nations Activities on Artificial Intelligence (AI)*, los progresos de la alianza *AI for Good* que lidera. La organización IEEE ha puesto en marcha una Iniciativa Mundial sobre la Ética de los Sistemas Inteligentes y Autónomos. La Unión Internacional de Telecomunicaciones y la Organización Mundial de la Salud han establecido también un Grupo de Debate sobre la "inteligencia artificial para la salud".

Las tecnologías de IA no son neutrales, sino que están intrínsecamente sesgadas por los datos en los que se basan y las decisiones que se toman durante la integración de esos datos (UNESCO, 2019a, apartado 3). En consecuencia, los debates actuales reflejan que hoy en día, en el plano mundial, sería necesario contar con una orientación ética universal global sobre los valores fundamentales que deben sustentar la elaboración de los sistemas de IA (UNESCO, Documento 2019a: apartado 11). En este sentido, la segunda premisa desde la que parte este estudio es que la mejor forma del instrumento a adoptar sería la de Recomendación. Los distintos informes mencionados en el estudio preliminar del Grupo de trabajo ampliado sobre la ética de la inteligencia artificial de la COMEST (Comisión Mundial de Ética del Conocimiento Científico y la Tecnología), coinciden en la conclusión de que no solo es deseable, sino asimismo urgente, que se adopten medidas para establecer un instrumento mundial no vinculante en forma de Recomendación. Una Recomendación, teniendo en cuenta su carácter no vinculante y el hecho de que

se centra en los principios y normas para la regulación internacional de una cuestión concreta, sería un método más flexible y adaptado a la complejidad de las cuestiones éticas planteadas por la IA (UNESCO, 2019a: apartado 14).

Recientemente, ya se ha señalado que la UNESCO aprobó el 23 de noviembre de 2021 la Recomendación sobre la ética de la Inteligencia Artificial (IA). Frente a otras opciones -como una Declaración o una propuesta de tratado internacional- desde la UNESCO se pensó que dicho instrumento podría constituir una herramienta fundamental para fomentar la elaboración de textos legislativos, políticas y estrategias nacionales e internacionales en el ámbito de la inteligencia artificial y reforzar su aplicación, así como para potenciar la cooperación internacional en torno al desarrollo y el uso éticos de la inteligencia artificial en apoyo de los Objetivos de Desarrollo Sostenible (ODS) (UNESCO, 2019b: punto 11). Coincide, de este modo, con la aproximación que desde Naciones Unidas se viene siguiendo con relación a la IA a través de cumbres anuales celebradas bajo el lema *AI for Good*. Liderada por la Unión Internacional de Telecomunicaciones, se trata de un foro en el que participan la mayoría de agencias y organismos de la ONU y en el que se busca promover un diálogo a nivel global sobre las potencialidades de la IA en conexión con los objetivos del desarrollo sostenible previstos en la Agenda 2030, siguiendo así la *Resolución de la Asamblea General de Naciones Unidas sobre el impacto del cambio tecnológico rápido en la consecución de los Objetivos de Desarrollo Sostenible y sus metas* y la estrategia del Secretario General de las Naciones Unidas en materia de nuevas tecnologías, de septiembre de 2018.

En dicha estrategia, el Secretario General de Naciones Unidas recordó la necesidad de colaborar estrechamente para superar los desafíos y conciliar los intereses, especialmente en las esferas de la privacidad y los derechos humanos, la ética, la igualdad y la equidad, la soberanía y la responsabilidad, la transparencia y la rendición de cuentas, a través de cinco principios: la protección y promoción de los valores globales establecidos en la Carta de las Naciones Unidas y la Declaración Universal de Derechos Humanos; el fomento de la inclusión y la transparencia entre gobiernos, empresas y sociedad civil para que se tomen decisiones colectivas con respecto a las nuevas tecnologías; el trabajo en colaboración estableciendo alianzas como el proyecto break through dentro del Pacto Mundial con el sector empresarial; el aprovechamiento de las capacidades y los mandatos existentes; finalmente, la humidad y el aprendizaje permanente.

Al hilo de esta cuestión, nos parece relevante mencionar el razonamiento seguido por el Grupo de trabajo de la COMEST al preparar su estudio preliminar sobre la ética de la inteligencia artificial para la Conferencia plenaria de la UNESCO. Tras valorar entre la disyuntiva de proponer una Resolución o

una Recomendación, concluyó que existía una gran heteronomía en los principios y en la implantación de los valores promovidos por distintos foros internacionales y regionales, debida no tanto a la definición que se hubiera elegido para la IA, como en razón de los objetivos que se persiguiesen a cada una de estas iniciativas: gobernanza, formación de los ingenieros, y política pública. La cuestión para el Grupo de trabajo, en consecuencia, fue la siguiente: ¿permitiría una Declaración de la UNESCO sobre la ética de la IA que esta heteronomía se atenuase en torno a unos pocos principios rectores que respondieran de manera exhaustiva a las cuestiones éticas de la IA, así como a las preocupaciones específicas de la UNESCO en los ámbitos de la educación, la cultura, la ciencia y la comunicación? El Grupo de Trabajo no estimó que esto fuera posible y percibió, incluso, el riesgo de que, durante el proceso de aprobación de dicha Declaración, los Estados pudiesen convenir esencialmente algunos principios generales, abstractos y no vinculantes, propios del formato de Declaración elegido. Desde esta perspectiva, se cuestionaron si aportaría la Declaración de la UNESCO sobre la ética de la IA un valor añadido frente a otras declaraciones e iniciativas en curso, concluyendo que nada hacía pensar que así fuera. En su opinión, podía ser cuestionable que el instrumento preparado por la UNESCO fuera a establecerse de manera inmediata como una referencia internacional, en un contexto de competencia entre marcos éticos, en un período en el que emergen distintas tecnologías y sus usos no se han estabilizado aún. Por lo tanto, el Grupo de trabajo consideró que una Recomendación constituía una herramienta más adecuada en la situación actual. (COMEST, 2019, apartados 99 a 101). Sobre todo, considerando -como comentamos más adelante en el epígrafe III- que la Recomendación aprobada en 2021 por la UNESCO sobre la ética de la IA incluye mecanismos de seguimiento y evaluación del grado de cumplimiento de la misma por parte de los 193 Estados miembros de esta Organización.

2. SIGNIFICADO DE LOS VALORES DE LA CARTA DE NACIONES UNIDAS EN EL DERECHO INTERNACIONAL

Hablar de valores en un sentido funcional (Abbagnano, 1982, 611) y, en concreto, de los valores de la Carta de Naciones Unidas como los principios rectores de esta Organización Internacional y de las relaciones que establecen entre sí sus Estados miembros, supone referirse a los objetivos que los redactores de este instrumento internacional pretendían alcanzar para el nuevo orden mundial surgido tras el final de la Segunda Guerra Mundial (los "propósitos" recogidos en el artículo primero de la Carta) y el modo de lograr su efectiva realización (los "principios" enunciados en el artículo segundo de dicho instrumento). De este modo, se estaba estableciendo la distinción que perdura hasta nuestros días, entre los valores instrumentales y los valores finales, permitiendo los primeros la concreción de los segundos. Entre los

valores instrumentales de la Carta de San Francisco pueden citarse: el arreglo pacífico de las controversias internacionales, la igualdad de los Estados y el corolario deber de éstos de no intervención en los asuntos internos de otros Estados, la cooperación y el deber de cumplir de buena fe las obligaciones que hayan libremente asumido. Por su parte, entre los valores finales destacan la paz, la seguridad internacional y la justicia.

Son valores que establecen obligaciones de todos (*omnium*) y frente a todos (*erga omnes*) pues, en última instancia, es la Humanidad la que se evoca como detentadora de los mismos cuando se afirma en el mismo preámbulo de la Carta: "Nosotros, los pueblos de las Naciones Unidas". Una idea, ésta, que ha sido reiterada en *la Declaración de los Principios del Derecho Internacional referentes a las relaciones de amistad y a la cooperación entre los Estados de conformidad con la Carta de las Naciones Unidas*, anexa a la Resolución 2625(XXV) que fue adoptada por la Asamblea General de esta Organización el 24 de octubre de 1970, la cual es hoy considerada expresión del Derecho Internacional general.

Los valores de la Carta de Naciones Unidas no sólo no han envejecido con el transcurso de los años sino que, por el contrario, son más fuertes y necesarios que nunca para regular todos los desafíos a los que ha de hacer frente la comunidad internacional en este siglo XXI. A esta realidad contribuye, sin duda, el hecho de que sean valores inclusivos, universales y atemporales, y que sean vistos como las reglas válidas para el juego de las relaciones internacionales en el que pueden participar los sujetos y actores internacionales presentes y futuros.

Son valores inclusivos en la medida en que hacen referencia a todos los Estados, no sólo a aquellos que vencieron la Segunda Guerra Mundial, de modo que han podido ser asumidos por los "perdedores" de aquella contienda y, además, ser compartidos por los nuevos Estados surgidos del proceso de descolonización iniciado en los años siguientes a su conclusión.

Son valores universales, a mayor abundamiento, debido al hecho de que no son tanto valores predicables del conjunto de Estados que integran la comunidad internacional, como de la propia comunidad internacional de Estados en su conjunto, en la que la idea del grupo sobresale sobre la de sus integrantes.

Son valores atemporales, finalmente, porque la idea de Humanidad incluye a las generaciones presentes y a las futuras, de manera que los valores de la Carta de Naciones Unidas son objeto de una constante relectura, como una puesta al día según las circunstancias y momento histórico en el que se reivindican.

3. RELEVANCIA DE LOS VALORES RECOGIDOS EN LA RECOMENDACIÓN DE LA UNESCO SOBRE LA ÉTICA DE LA INTELIGENCIA ARTIFICIAL (IA)

La inteligencia artificial tiene importantes implicaciones sociales y culturales. Al igual que otras muchas tecnologías de la información, la IA plantea cuestiones relacionadas con la libertad de expresión, la privacidad y la vigilancia, la propiedad de los datos, el sesgo y la discriminación, manipulación de la información y la confianza, las relaciones de poder, y el impacto medioambiental en lo que se refiere a su consumo de energía. Además, la IA plantea específicamente nuevos retos relacionados con su interacción con las capacidades cognitivas humanas. Los sistemas basados en la IA tienen consecuencias para la comprensión y la experiencia humanas. Los algoritmos de las redes sociales y los sitios de noticias pueden facilitar la propagación de desinformación y repercutir en el significado percibido de los "hechos" y la "verdad", así como en la interacción y la participación en el ámbito político. El aprendizaje automático puede integrar y exacerbar el sesgo, lo que puede dar lugar a la desigualdad, a la exclusión, y a una amenaza a la diversidad cultural. La escala y el poder generados por la tecnología de la IA acentúa la asimetría entre individuos, grupos y países incluida la denominada "brecha digital" en cada nación, y entre naciones. Tal brecha puede verse agravada por la falta de acceso a elementos fundamentales como los algoritmos para el aprendizaje y la clasificación, los datos para capacitar y evaluar los algoritmos, los recursos humanos para codificar y configurar el software y preparar los datos, así como los recursos computacionales para el almacenamiento y el procesamiento de los datos (COMEST, 2019: apartado 22).

A la luz de estas consideraciones, resulta especialmente relevante que la UNESCO declare, en el parágrafo quinto del Preámbulo de su Recomendación de 2021 sobre la ética de la IA, que se halla "guiada por los propósitos y principios de la Carta de las Naciones Unidas". A mayor abundamiento, en el parágrafo 13 de este mismo Preámbulo se observa que el marco normativo para las tecnologías de la IA y sus implicaciones sociales se fundamenta en los marcos jurídicos internacionales y nacionales, los derechos humanos y las libertades fundamentales, la ética (…) y conecta los valores y principios éticos con los retos y oportunidades vinculados a las tecnologías de la IA, sobre la base de un entendimiento común y unos objetivos compartidos." Añadiendo en el parágrafo siguiente (el décimo cuarto) que "los valores y principios éticos pueden ayudar a elaborar y aplicar medidas de política y normas jurídicas basadas en los derechos, proporcionando orientación con miras al rápido desarrollo tecnológico."

En su estudio preliminar, el Grupo de Trabajo de la COMEST al que me referí páginas atrás, propuso una serie de principios genéricos para el desarrollo, la implantación y el uso de la IA que no han sido reproducidos en su tenor literal, en la Recomendación aprobada por la Conferencia Plenaria de la UNESCO en 2021 sobre la ética de la IA. Estos principios eran los siguientes:

a) Derechos humanos: la IA debe desarrollarse e implementarse de acuerdo con las normas internacionales de los derechos humanos.

b) Integración: la IA debe ser inclusiva, con el objetivo de evitar sesgos, propiciar la diversidad y prevenir una nueva brecha digital.

c) Prosperidad: la IA debe desarrollarse para mejorar la calidad de vida.

d) Autonomía: la IA debe respetar la autonomía humana mediante la exigencia del control humano en todo momento.

e) Explicabilidad: la IA debe ser explicable, capaz de proporcionar una idea de su funcionamiento.

f) Transparencia: los datos utilizados para capacitar los sistemas de IA deben ser transparentes.

g) Conocimiento y capacitación: el conocimiento de los algoritmos y una comprensión básica del funcionamiento de la IA son necesarios para capacitar a los ciudadanos.

h) Responsabilidad: los desarrolladores y las empresas deben tener en cuenta la ética al desarrollar los sistemas inteligentes autónomos.

i) Asunción de responsabilidades: Deben desarrollarse mecanismos que permitan atribuir responsabilidades respecto a las decisiones basadas en la IA y la conducta de los sistemas de IA. j) Democracia: la IA debe desarrollarse, implantarse y utilizarse con arreglo a principios democráticos.

k) Buena gobernanza: los gobiernos deben presentar informes periódicos sobre su utilización de la IA en los ámbitos de la actividad policial, la inteligencia y la seguridad.

l) Sostenibilidad: En todas las aplicaciones de la AI, los beneficios potenciales deben equilibrarse con el impacto medioambiental del ciclo de producción completo de la IA y las tecnologías de la información (COMEST, 2019, apartado 107).

A pesar de esta propuesta, los Estados participantes en la Asamblea General de la UNESCO decidieron centrarse en aquellos valores que podían conectarse con la Carta de Naciones Unidas, lo cual es ciertamente relevante, si se considera que la Recomendación aprobada en noviembre de 2021 por la

UNESCO incluye mecanismos de seguimiento y evaluación de las políticas, los programas y los mecanismos relativos a la ética de la IA que los Estados adopten -de acuerdo con sus circunstancias- de forma creíble y transparente (UNESCO, 2021, apartado 131). En este sentido, se presentan varios posibles mecanismos de seguimiento y evaluación, tales como una comisión de ética, un observatorio de ética de la IA, un repositorio que abarque el desarrollo ético y conforme a los derechos humanos de los sistemas de IA, o contribuciones a las iniciativas existentes para reforzar la conformidad a los principios éticos en todas las esferas de competencia de la UNESCO, un mecanismo de intercambio de experiencias, entornos de pruebas reguladores de la IA y una guía de evaluación para que todos los actores de la IA determinen en qué medida cumplen las recomendaciones de actuación mencionadas en esta Recomendación (UNESCO, 2021, apartado 134).

Tales procesos de seguimiento y evaluación deberían asegurar una amplia participación de todas las partes interesadas, entre ellas, aunque no exclusivamente, las personas vulnerables o en situación de vulnerabilidad (UNESCO 2021: apartado 132). Se añade además que se debería garantizar la diversidad social, cultural y de género, con miras a mejorar los procesos de aprendizaje y fortalecer los nexos entre las conclusiones, la adopción de decisiones, la transparencia y la rendición de cuentas sobre los resultados.

En este mismo sentido, los Estados deberían elaborar instrumentos e indicadores adecuados para evaluar su eficacia y eficiencia en función de normas, prioridades y objetivos acordados, incluidos objetivos específicos para las personas pertenecientes a poblaciones desfavorecidas y marginadas y personas vulnerables o en situación de vulnerabilidad, así como el impacto de los sistemas de IA en los planos individual y social. El seguimiento y la evaluación del impacto de los sistemas de IA y de las políticas y prácticas conexas relativas a la ética de la IA deberían realizarse de manera continua y sistemática proporcionalmente a los riesgos correspondientes. Este proceso debería basarse en marcos acordados internacionalmente e ir acompañado de evaluaciones de instituciones, proveedores y programas públicos y privados, incluidas autoevaluaciones, así como de estudios de seguimiento y la elaboración de conjuntos de indicadores. La recopilación y el procesamiento de datos deberían realizarse de conformidad con el Derecho internacional, la legislación nacional en materia de protección y confidencialidad de datos y los valores y principios enunciados en la Recomendación de la UNESCO sobre la ética de la IA (UNESCO, 2021, apartado 133).

4. EXAMEN CRÍTICO DE LOS VALORES COMPARTIDOS CON LA CARTA DE NACIONES UNIDAS: LA JUSTICIA SOCIAL EN LAS CONSIDERACIONES ÉTICAS DE LA INTELIGENCIA ARTIFICIAL

La referencia a la idea de justicia social que subyace al principio ético de hacer una IA en beneficio de toda la Humanidad, a través de la solidaridad entre los países del mundo, es una vieja reivindicación del Derecho internacional, desde que se proclamó la Declaración Universal de los Derechos Humanos por la Asamblea General de Naciones Unidas, como anexa a la Resolución 217 (III), aprobada sin oposición el 10 de diciembre de 1948.

En el párrafo segundo del Preámbulo de este referente mundial en la protección de los derechos humanos se afirma lo siguiente: "Considerando que… se ha proclamado como la aspiración más elevada del hombre el advenimiento de un mundo en el que los seres humanos, liberados del temor y de la miseria, disfruten de la libertad de palabra y de la libertad de creencias." Es importante evocar estas referencias a la miseria y a un concepto más amplio de libertad, dentro del Preámbulo de la Declaración Universal de los Derechos Humanos porque es una contribución del mundo occidental -enfrentado durante la Guerra Fría a su antagónico socialista- pese a lo que erróneamente pudiera pensarse tras hacer bandera la Unión Soviética y sus satélites, de los derechos sociales y económicos, frente a la defensa de los derechos civiles y políticos por parte de Estados Unidos y sus aliados.

En efecto, un breve ejercicio de memoria histórica es necesario para hacer frente a las voces que pretenden desacreditar las aspiraciones de justicia social y de solidaridad en el aprovechamiento de las potencialidades de la IA como quimeras y ensoñaciones de países en desarrollo que confunden -así se les critica- sus deseos y aspiraciones con la realidad del mundo en el que vivimos.

Fue el presidente de Estados Unidos, Franklin D. Roosevelt, quien el 6 de enero de 1941 proclamó un discurso ante el Congreso norteamericano en el que evocaba cuatro libertades fundamentales de cualquier ser humano: la libertad de palabra y de pensamiento; la libertad religiosa; la libertad ante la necesidad (esto es, el germen de la idea de justicia social y de los derechos económicos y sociales); y la libertad frente al miedo, o lo que es igual, la obligación moral del desarme dirigido a prevenir agresiones armadas entre los Estados (Cassese, 1991, 37).

La evocación de la idea de justicia social por parte del presidente que lideró el *New Deal* en su país, defendiendo el rescate social y económico de todos los americanos como motor de desarrollo de la nación, no fue aprovechado por el resto de países occidentales que durante años pusieron su acento sólo sobre los derechos civiles y políticos (las dos primeras libertades

enunciadas por Roosevelt en su discurso), en detrimento de los derechos económicos y sociales, conectados con la libertad ante la necesidad.

Esta actitud fue un retroceso reprochable, en mi opinión, con respecto a los avances que supuso, al concluir la Primera Guerra Mundial, la inclusión del término "justicia" en el Preámbulo del Pacto de la Sociedad de Naciones, recogido en la Parte Primera del Tratado de Paz de Versalles, firmado el 28 de junio de 1919: "Las Altas Partes contratantes, considerando que para fomentar la cooperación entre las naciones y para garantizar la paz y la seguridad importa… mantener a la luz del día relaciones internacionales fundadas sobre la justicia… hacer que reine la justicia." (Carrillo Salcedo, 1985, 235).

Los vencedores de la Primera Guerra Mundial habían comprendido que determinadas causas económicas y sociales podían llevar a la guerra, por lo que era necesario crear condiciones sociales y económicas de paz (Carrillo Salcedo, 2001, 40). De ahí que crearan la Organización Internacional del Trabajo en la Parte XIII del Tratado de Paz de Versalles, al tiempo que incluían una referencia expresa a la justicia social en el Preámbulo de dicha Parte XIII: "Considerando que existen condiciones de trabajo que implican para un gran número de personas la injusticia, la miseria y las privaciones, lo cual engendra tal descontento que la paz y las armonías universales peligran" (párrafo segundo). "Las Altas Partes contratantes, movidas por sentimientos de justicia y de humanidad, así como por el deseo de asegurar una paz mundial…" (párrafo cuarto); y otra mención indirecta en el artículo 23 del Pacto de la Sociedad de Naciones: "Con la reserva y de conformidad con las disposiciones de los convenios internacionales existentes en la actualidad o que se celebren en lo sucesivo, los Miembros de la Sociedad: a) se esforzarán en asegurar y mantener condiciones de trabajo equitativas y humanitarias para el hombre, la mujer y el niño en sus propios territorios, así como en todos los países a que se extienden sus relaciones de comercio y de industria y, para este fin, fundarán y mantendrán las necesarias organizaciones internacionales."

En este sentido, el Profesor Carrillo Salcedo ha señalado que la Sociedad de Naciones no se limitó a los aspectos políticos de la paz -ausencia de conflicto- sino que consideró que en el mantenimiento de la paz entre los Estados entraban en consideración otras dimensiones, tales como la toma de conciencia de las causas económicas y sociales que podían llevar a la guerra. Era necesario crear condiciones económicas y sociales de paz de modo que, en palabras de este autor, los Estados fundadores de la Sociedad de Naciones consideraban que la paz universal sólo podía basarse en la justicia social y que, prevaleciendo la injusticia social en un gran número de países, se estaba poniendo en peligro la paz mundial (Carrillo Salcedo, 1985, 47).

La idea de justicia social que inspiró a los vencedores de la Primera Guerra Mundial se aprecia también en la Carta de San Francisco, de 26 de junio de 1945, constitutiva de las Naciones Unidas, si bien, de una manera más matizada, en su preámbulo: "Nosotros, los pueblos de las Naciones Unidas resueltos… a crear condiciones bajo las cuales pueda manifestarse la justicia…" Pero, sobre todo, en su artículo primero, en el que se enuncia la justicia como un principio instrumental del mantenimiento de la paz y la seguridad internacionales. Aun cuando falta el calificativo "social" a la idea de justicia en este artículo, a diferencia del Pacto de la Sociedad de Naciones, creo que los redactores de la Carta de Naciones Unidas compartían la misma convicción de los creadores de la Sociedad de Naciones, en el sentido de que la paz duradera entre los Estados debía basarse sobre la idea de justicia en todas sus dimensiones, incluida la social.

La idea de justicia social es objeto de específica consideración, dentro del Preámbulo de la Recomendación de la UNESCO (parágrafos 7 y 8) cuando se afirma que: "Reconociendo también que las tecnologías de la IA pueden agravar las divisiones y desigualdades existentes en el mundo, dentro de los países y entre ellos, y que *es preciso defender la justicia, la confianza y la equidad para que ningún país y ninguna persona se queden atrás*, ya sea mediante el acceso equitativo a las tecnologías de la IA y el disfrute de los beneficios que aportan o mediante la protección contra sus consecuencias negativas, reconociendo al mismo tiempo las diferentes circunstancias de los distintos países y respetando el deseo de algunas personas de no participar en todos los avances tecnológicos." (La cursiva es añadida). "Consciente de que todos los países se enfrentan a una aceleración del uso de las tecnologías de la información y la comunicación y las tecnologías de la IA, así como a una necesidad cada vez mayor de alfabetización mediática e informacional, y de que la economía digital presenta importantes desafíos sociales, económicos y ambientales y ofrece oportunidades de compartir los beneficios, especialmente para los países de ingreso mediano bajo, incluidos, entre otros, los países menos adelantados (PMA), los países en desarrollo sin litoral (PDSL) y los pequeños Estados insulares en desarrollo (PEID), que requieren el reconocimiento, la protección y la promoción de las culturas, los valores y los conocimientos endógenos a fin de desarrollar economías digitales sostenibles."

La idea de justicia social -en un sentido similar al apreciado en el Derecho internacional representado por la Carta de Naciones Unidas- está evocada en otras dos ocasiones, parágrafos 18 y 21 del Preámbulo de la Recomendación de la Unesco, pero sobre todo en el cuerpo normativo de este documento que desarrolla los valores (UNESCO, 2021, apartados 22 a 24) y los principios éticos de la IA identificados (UNESCO, 2021, apartados 28 a 30) y que establecen los ámbitos de actuación política a través de los cuales se ponen en práctica dichos

valores y principios. De hecho, es particularmente llamativo el descubrir que en los once ámbitos de actuación contemplados en esta Recomendación, la idea de justicia social aparece como un elemento transversal a todos ellos, de una manera más o menos explícita.

Así, entre los valores identificados para una ética de la IA en la Recomendación de la UNESCO, se incluye *vivir en sociedades pacíficas, justas e interconectadas*, para lo cual, los actores de la IA deberían propiciar sociedades pacíficas y justas, sobre la base de un futuro interconectado en beneficio de todos, compatibles con los derechos humanos y las libertades fundamentales, y participar en su construcción (UNESCO, 2021, apartado 22). Este valor exige que se promuevan la paz, la inclusión y la justicia, la equidad y la interconexión durante el ciclo de vida de los sistemas de IA, en la medida en que los procesos de dicho ciclo de vida no deberían segregar ni cosificar a los seres humanos y las comunidades ni mermar su libertad, su autonomía de decisión y su seguridad, así como tampoco dividir y enfrentar entre sí a las personas y los grupos ni amenazar la coexistencia entre los seres humanos, los demás seres vivos y el medio natural (UNESCO, 2021, apartado 24).

Por su parte, entre los principios proclamados en esta Recomendación, expresamente se incluye *la equidad y no discriminación* y a tal fin se afirma (UNESCO, 2021, apartado 28) que los actores de la IA deberían promover la justicia social, salvaguardar la equidad y luchar contra todo tipo de discriminación, de conformidad con el Derecho internacional. Ello supone adoptar un enfoque inclusivo para garantizar que los beneficios de las tecnologías de la IA estén disponibles y sean accesibles para todos, teniendo en cuenta las necesidades específicas de los diferentes grupos de edad, los sistemas culturales, los diferentes grupos lingüísticos, las personas con discapacidad, las niñas y las mujeres y las personas desfavorecidas, marginadas y vulnerables o en situación de vulnerabilidad. Asimismo, los Estados deberían esforzarse por promover un acceso inclusivo para todos, incluidas las comunidades locales, a sistemas de IA con contenidos y servicios adaptados al contexto local, y respetando el multilingüismo y la diversidad cultural. Los Estados deberían, igualmente, esforzarse por reducir las brechas digitales y garantizar el acceso inclusivo al desarrollo de la IA y la participación en él. En el plano nacional, los Estados deberían promover la equidad entre las zonas rurales y urbanas y entre todas las personas, con independencia de su raza, color, ascendencia, género, edad, idioma, religión, opiniones políticas, origen nacional, étnico o social, condición económica o social de nacimiento, discapacidad o cualquier otro motivo, en lo que respecta al acceso al ciclo de vida de los sistemas de IA y la participación en él.

En el plano internacional, los países más avanzados tecnológicamente tienen la responsabilidad de ser solidarios con los menos avanzados para garantizar que los beneficios de las tecnologías de la IA se compartan de manera que, para estos últimos, el acceso al ciclo de vida de los sistemas de IA y la participación en él contribuyan a un orden mundial más equitativo en lo que respecta a la información, la comunicación, la cultura, la educación, la investigación y la estabilidad socioeconómica y política (UNESCO, 2021, apartado 28).

Como ya se ha indicado, una de las características más sobresalientes de la Recomendación de la UNESCO sobre la ética de la IA es la previsión de ámbitos de acción política para los Estados. Esto es, se busca que los valores y principios proclamados sean puestos en práctica de manera efectiva. En este sentido, se indica en su apartado 48 que la principal acción consiste en que los Estados Miembros establezcan medidas eficaces, por ejemplo, marcos o mecanismos normativos, y velen por que otras partes interesadas, como las empresas del sector privado, las instituciones universitarias y de investigación y la sociedad civil, se adhieran a ellas, sobre todo alentando a todas las partes interesadas a que elaboren instrumentos de evaluación del impacto en los derechos humanos, el estado de derecho, la democracia y la ética, así como instrumentos de diligencia debida, de conformidad con las orientaciones, incluidos los Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas. El proceso de elaboración de esas políticas o mecanismos debería incluir a todas las partes interesadas y tener en cuenta las circunstancias y prioridades de cada Estado Miembro.

Nos llama poderosamente la atención descubrir que en los once ámbitos de actuación contemplados en la Recomendación de la UNESCO sobre la ética en la IA (1. Evaluación del impacto ético; 2. Gobernanza y administración éticas; 3. Política de datos; 4. Desarrollo y cooperación internacional; 5. Medioambiente y ecosistemas; 6. Género; 7. Cultura; 8. Educación e investigación; 9. Comunicación e información; 10. Economía y trabajo; y 11. Salud y bienestar social), la idea de justicia social aparece como un elemento transversal a todos ellos, de una manera más o menos explícita, lo que evidencia que, de entre todos los principios y valores proclamados en este instrumento, éste sería sino el más importante, sí al menos, uno de los más relevantes y sin duda, ubicado en el núcleo duro de los valores que deben presidir la ética aplicada a la IA.

Así, tratándose del primer ámbito de actuación política previsto en la Recomendación de la UNESCO, la *evaluación del impacto ético,* se defiende la necesidad de que los Estados evalúen los efectos socioeconómicos de los sistemas de IA en la pobreza y velar por que la brecha entre los ricos y los pobres, así como la brecha digital entre los países y dentro de ellos, no

aumenten con la adopción masiva de tecnologías de la IA en la actualidad y en el futuro. Para ello, en particular, deberían aplicarse protocolos de transparencia ejecutables, que correspondan al acceso a la información, incluida la información de interés público en poder de entidades privadas (UNESCO 2021: apartado 51).

En lo que respecta al segundo ámbito de actuación política, la *gobernanza y administración éticas,* los Estados deberían fomentar el desarrollo y la accesibilidad de un ecosistema digital para el desarrollo ético e inclusivo de los sistemas de IA en el plano nacional, en particular con miras a reducir las diferencias de acceso durante el ciclo de vida de los sistemas de IA, contribuyendo al mismo tiempo a la colaboración internacional. Ese ecosistema incluiría, en particular, tecnologías e infraestructuras digitales y mecanismos para compartir los conocimientos en materia de IA, según proceda. Además, los Estados deberían establecer mecanismos, en colaboración con las organizaciones internacionales, las empresas transnacionales, las instituciones universitarias y la sociedad civil, para garantizar la participación activa de todos los Estados, especialmente los países de ingreso mediano bajo, en particular los países menos adelantados (PMA), los países en desarrollo sin litoral (PDSL) y los pequeños Estados insulares en desarrollo (PEID), en los debates internacionales sobre la gobernanza de la IA. Esto puede hacerse mediante la provisión de fondos, garantizando la participación regional en condiciones de igualdad, o mediante cualquier otro mecanismo. Además, para velar por que los foros sobre la IA sean inclusivos, los Estados deberían facilitar los desplazamientos de los actores de la IA dentro y fuera de su territorio, especialmente los de los países de ingreso mediano bajo, en particular los PMA, los PDSL y los PEID, para que puedan participar en esos foros (UNESCO, 2021, apartados 59 y 60).

En relación con el tercer ámbito de actuación política, la *política de datos,* es de destacar la apuesta por los datos abiertos y por el patrimonio digital común. En cuanto a lo primero, los Estados deberían promover los datos abiertos y promover mecanismos, como repositorios abiertos de datos y códigos fuente públicos o de financiación pública y fideicomisos de datos, a fin de apoyar el intercambio seguro, equitativo, legal y ético de datos, entre otros (UNESCO 2021: apartado 75). Respecto de la segunda medida, los Estados, con el apoyo de las Naciones Unidas y la UNESCO, deberían adoptar un enfoque de patrimonio digital común respecto a los datos, cuando proceda, aumentar la interoperabilidad de los instrumentos y conjuntos de datos, así como las interfaces de los sistemas que albergan datos, y alentar a las empresas del sector privado a que compartan con todas las partes interesadas los datos que recopilan, en beneficio de la investigación, la innovación o el interés público, según proceda (UNESCO, 2021, apartado 77).

El cuarto ámbito contemplado para hacer efectivos los valores y principios enunciados en la Recomendación de la UNESCO sobre la ética de la IA, *el desarrollo y la cooperación internacionales*, contempla varias medidas: los Estados deberían velar por que la utilización de la IA en esferas relacionadas con el desarrollo, como la educación, la ciencia, la cultura, la comunicación y la información, la atención sanitaria, la agricultura y el suministro de alimentos, el medio ambiente, la gestión de recursos naturales y de infraestructuras y la planificación y el crecimiento económicos, entre otras, respete los valores y principios enunciados en la presente Recomendación (UNESCO, 2021, apartado 79). Asimismo, se afirma que los Estados deberían procurar, por conducto de organizaciones internacionales, establecer plataformas de cooperación internacional en el ámbito de la IA para el desarrollo, en particular aportando competencias técnicas, financiación, datos, conocimientos del sector e infraestructura y facilitando la colaboración entre múltiples partes interesadas para hacer frente a los problemas complejos en materia de desarrollo, especialmente para los países de ingreso mediano bajo, en particular los PMA, los PDSL y los PEID (UNESCO, 2021, apartado 80). Igualmente, los Estados deberían procurar promover la colaboración internacional en materia de investigación e innovación en IA, especialmente en centros y redes de investigación e innovación que promuevan una mayor participación y liderazgo de los investigadores procedentes de países de ingreso mediano bajo y otros países, en particular de PMA, PDSL y PEID (UNESCO, 2021, apartado 81).

El Medioambiente y los ecosistemas es el quinto ámbito de actuación política enunciado en la Recomendación de la UNESCO. Al respecto, los Estados deberían establecer incentivos, cuando sea necesario y apropiado, para garantizar la elaboración y adopción de soluciones basadas en los derechos y en la ética de la IA en favor de la resiliencia ante el riesgo de desastres; la vigilancia, protección y regeneración del medio ambiente y los ecosistemas; y la preservación del planeta. Esos sistemas de IA deberían contar, durante todo su ciclo de vida, con la participación de las comunidades locales e indígenas y apoyar enfoques del tipo de economía circular y modalidades de consumo y producción sostenibles (UNESCO 2021: apartado 85).

Tratándose del sexto ámbito de actuación previsto, *la perspectiva de género*, la referencia es menos explícita que en otras ocasiones, cuando se establece que los Estados deberían velar por que se aproveche el potencial de los sistemas de IA para impulsar el logro de la igualdad de género. Deberían asegurarse de que estas tecnologías no exacerben las ya amplias brechas que existen entre los géneros en varios ámbitos del mundo analógico, sino que, al contrario, las eliminen. Entre estas brechas cabe citar la disparidad salarial entre hombres y mujeres; su representación desigual en ciertas profesiones y actividades; la falta de representación en los puestos directivos superiores, las juntas directivas o los

equipos de investigación en el campo de la IA; la brecha educativa; las desigualdades en el acceso, la adopción, la utilización y la asequibilidad de la tecnología digital y de la IA; y la distribución desigual del trabajo no remunerado y de las responsabilidades de cuidado en nuestras sociedades (UNESCO, 2021, apartado 89).

En lo que respecta al séptimo ámbito de actuación política, la *cultura*, los Estados deberían promover el conocimiento y la evaluación de los instrumentos de IA entre las industrias culturales locales y las pequeñas y medianas empresas que trabajan en el ámbito de la cultura, a fin de evitar el riesgo de concentración en el mercado cultural (UNESCO, 2021, apartado 97).

En relación con el octavo ámbito de actuación, *la educación y la investigación*, los Estados deberían colaborar con organizaciones internacionales, instituciones educativas y entidades privadas y no gubernamentales para impartir al público de todos los países, a todos los niveles, conocimientos adecuados en materia de IA, a fin de empoderar a la población y reducir las brechas digitales y las desigualdades en el acceso a la tecnología digital resultantes de la adopción a gran escala de sistemas de IA (UNESCO, 2021, apartado 101). Además, los Estados deberían promover la participación y el liderazgo de las niñas y las mujeres, las personas de diversos orígenes étnicos y culturas, las personas con discapacidad, las personas marginadas y vulnerables o en situación de vulnerabilidad y las minorías, así como de todas aquellas personas que no gocen plenamente de los beneficios de la inclusión digital, en los programas de educación en materia de IA en todos los niveles, así como el seguimiento y el intercambio con otros Estados de las mejores prácticas en este ámbito (UNESCO, 2021, apartado 105).

El noveno ámbito de actuación es *la comunicación y la información* y al respecto, se acuerda que los Estados deberían invertir en competencias digitales y de alfabetización mediática e informacional y promoverlas, a fin de reforzar el pensamiento crítico y las competencias necesarias para comprender el uso y las implicaciones de los sistemas de IA, con miras a atenuar y contrarrestar la desinformación, la información errónea y el discurso de odio. Una mejor comprensión y evaluación de los efectos tanto positivos como potencialmente perjudiciales de los sistemas de recomendación debería formar parte de esos esfuerzos (UNESCO, 2021, apartado 114).

En relación con el décimo ámbito de actuación, relativos a *la economía y al trabajo,* se contempla que los Estados deberían evaluar y abordar el impacto de los sistemas de IA en los mercados de trabajo y sus consecuencias en las necesidades educativas en todos los países y, más concretamente, en los países cuya economía requiere mucha mano de obra. Para ello puede ser preciso introducir una gama más amplia de competencias "básicas" e

interdisciplinarias en todos los niveles educativos, a fin de dar a los trabajadores actuales y a las nuevas generaciones una oportunidad equitativa de encontrar empleo en un mercado en rápida evolución y para asegurar que sean conscientes de los aspectos éticos de los sistemas de IA (UNESCO, 2021, apartado 116). Asimismo, los Estados deberían adoptar las medidas adecuadas para garantizar la competitividad de los mercados y la protección de los consumidores, considerando posibles medidas y mecanismos en los planos nacional, regional e internacional, a fin de impedir los abusos de posición dominante en el mercado, incluidos los monopolios, en relación con los sistemas de IA durante su ciclo de vida, ya se trate de datos, investigación, tecnología o mercados. Los Estados deberían prevenir las desigualdades resultantes, evaluar los mercados correspondientes y promover mercados competitivos. Se debería prestar la debida atención a los países de ingreso mediano bajo, en particular a los PMA, los PDSL y los PEID, que están más expuestos y son más vulnerables a la posibilidad de que se produzcan abusos de posición dominante en el mercado, como consecuencia de la falta de infraestructuras, capacidad humana y reglamentación, entre otros factores. Los actores de la IA que desarrollen sistemas de IA en países que hayan establecido o adoptado normas éticas en materia de IA deberían respetar estas normas cuando exporten estos productos, desarrollen sus sistemas de IA o los apliquen en países donde no existan dichas normas, respetando al mismo tiempo el Derecho internacional y las leyes, normas y prácticas nacionales aplicables de estos países (UNESCO, 2021, apartado 120).

Finalmente, en conexión con el undécimo y último ámbito de actuación, *la salud y el bienestar social,* se prevé que los Estados deberían esforzarse por emplear sistemas eficaces de IA para mejorar la salud humana y proteger el derecho a la vida, en particular atenuando los brotes de enfermedades, al tiempo que desarrollan y mantienen la solidaridad internacional para hacer frente a los riesgos e incertidumbres relacionados con la salud en el plano mundial, y garantizar que su despliegue de sistemas de IA en el ámbito de la atención de la salud sea conforme al Derecho internacional y a sus obligaciones en materia de derechos humanos. Los Estados deberían velar por que los actores que participan en los sistemas de IA relacionados con la atención de la salud tengan en cuenta la importancia de las relaciones del paciente con su familia y con el personal sanitario (UNESCO, 2021, apartado 121).

5. CONCLUSIONES

Al finalizar estas páginas es posible resumir los planteamientos anteriormente defendidos en las siguientes ideas a modo de síntesis: la IA es un medio y no un fin en sí misma. El fin es el desarrollo de los pueblos, de conformidad con los valores compartidos de la Carta de Naciones y de los

instrumentos internacionales que reivindican una aproximación ética al ciclo de vida de los sistemas de IA. Sin el desarrollo de los pueblos no puede afianzarse la paz, la seguridad y la justicia internacionales. Así lo entendieron nuestros padres y abuelos que desarrollaron un papel en la creación de la Organización de Naciones Unidas y, antes de ella, de la Sociedad de Naciones, tras sufrir el flagelo de dos Guerras Mundiales. Por ello es tan importante no perder de referencia que la IA es un medio y no un fin en sí misma.

Como corolario de la idea anterior, hay que recordar que algunos fines no pueden alcanzarse por cualquier medio pues, aun siendo en sí legítimos tales fines, estos se deslegitiman si se consiguen a cualquier precio. De ahí que se defienda desde hace años a todos los niveles unos principios éticos para la IA en los que el respeto de la dignidad humana y la protección de los derechos humanos fundamentales, sea su piedra angular.

En esta misión de asegurar el desarrollo de los pueblos respetando los derechos humanos fundamentales están implicados todos los sujetos y actores internacionales que participan en alguna medida en el ciclo de vida de los sistemas de IA. Nadie queda fuera ni puede pretender eludir su responsabilidad a este respecto. Nos jugamos demasiado en ello.

6. BIBLIOGRAFÍA

Abbagnano, Nicola (1982), *Diccionario de Filosofía,* Fondo de Cultura Económica, México-Buenos Aires.

Boden, Margaret (2016), *AI: Its Nature and Future*, Oxford University Press, Oxford.

Carrillo Salcedo, Juan Antonio (1985), *El Derecho Internacional en un mundo en cambio*, Tecnos, Madrid.

— (2001), *Soberanía de los Estados y Derechos Humanos,* Tecnos, Madrid.

Cassese, Antonio (1991), *Los derechos humanos en el mundo contemporáneo*, Ariel, Barcelona.

COMEST (Comisión Mundial de Ética del Conocimiento Científico y la Tecnología), "Estudio preliminar sobre la ética de la inteligencia artificial" SHS/COMEST/EXTWG-ETHICS-AI/2019/1, de 26 de febrero de 2019. Disponible en: https://unesdoc.unesco.org/ark:/48223/pf0000367823 Visitado el 1 de marzo de 2022.

Consejo de Europa (2020), *Recomendación CM/Rec (2020) 1 del Comité de Ministros a los Estados Miembros sobre los impactos de los sistemas algorítmicos en los derechos humanos*, el 8 de abril de 2020, Disponible en https://rm.coe.int/09000016809e1154 (visitado el 1 de marzo de 2022)

García San José, Daniel (2021a), "Implicaciones jurídicas y bioéticas de la inteligencia artificial (IA). Especial consideración al marco normativo internacional", en: *Cuadernos de Derecho Transnacional*, 13, 1, 255-276.

— (2021b), "International lawyers' contribution to a friendly artificial intelligence", en: Llano Alonso, Fernando, Joaquín Garrido Martín (eds.), *Inteligencia artificial y Derecho. El jurista ante los retos de la era digital*, Thomson-Reuter Aranzadi, Cizur Menor, 133-152.

Institute of Electrical and Electronic Engineers (IEEE) (2018), *Ethically Aligned Design* Version 2 for Public Discussion, New Jersey, disponible en: https://ethicsinaction.ieee.org/ Visitado el 1 de marzo de 2022.

Naciones Unidas, Estrategia del Secretario General de las Naciones Unidas en materia de nuevas tecnologías, de septiembre de 2018, disponible en https://www.un.org/en/newtechnologies/ Visitado el 1 de marzo de 2022.

— (2018), Resolución de la Asamblea General A/RES/73/17 de 3 de diciembre de 2018. Disponible en internet en https://undocs.org/pdf?symbol=es/A/RES/73/17 Visitado el 1 de marzo de 2022.

— (2022), Proyecto Breakthrough (IA para todos) de Naciones Unidas https://breakthrough.unglobalcompact.org/site/assets/files/1454/hhw-16-0017-d_c_artificial_intelligence.pdf Visitado el 1 de marzo de 2022.

OCDE (2019), Principios Éticos de la OCDE en materia de Inteligencia Artificial, *op. cit.*, adoptados en mayo de 2019 y posteriormente incluidos como anexo a la Declaración Ministerial sobre Comercio y Economía Digital, del G20 en su cumbre de Tsukuba, Japón, en junio de ese mismo año.

OECD (2019), Going Digital. Paris, disponible en: http://www.oecd.org/going-digital/ai/ Visitado el 1 de marzo de 2022.

UNESCO (2019a), Documento 206 EX/42, Estudio preliminar sobre los aspectos técnicos y jurídicos relativos a la conveniencia de disponer de un instrumento normativo sobre la ética de la inteligencia artificial, aprobado por la Conferencia General de la Organización el 30 de julio de 2019. Disponible en: https://unesdoc.unesco.org/ark:/48223/pf0000369455_spa Visitado el 1 de marzo de 2022.

— (2019b), Estudio preliminar sobre un posible instrumento normativo relativo a la ética de la inteligencia artificial, Doc. 40 C/67, aprobado por la Conferencia General de la Organización el 30 de julio de 2019. Disponible en: https://unesdoc.unesco.org/ark:/48223/pf0000369455_spa Visitado el 1 de marzo de 2022.

— (2020) Anteproyecto de Recomendación sobre la Ética de la IA, SHS/BIO/AHEG-AI/2020/4 REV 2, Doc. Final, Paris, 7 de septiembre de 2020. Disponible en: https://unesdoc.unesco.org/search/N-EXPLORE-47510aaf-a4ce-45d0-8296-791ecfefa7c8 Visitado el 1 de marzo de 2022.

— (2021) Recomendación sobre la Ética de la IA, aprobada el 23 de noviembre de 2021. Disponible en https://unesdoc/unesco.org/ark:/48223/pf0000380455_spa Visitado el 1 de marzo de 2022.

Unión Europea (2019), Directrices para una Inteligencia Artificial fiable del Grupo de Expertos de Alto Nivel creado por la Comisión Europea y publicadas en 2019 Disponible en: https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines≠/ Visitado el 1 de marzo de 2022.

— (2020), Libro Blanco de la Comisión Europea sobre la Inteligencia Artificial: un enfoque europea orientado hacia la excelencia y la confianza, COM (2020) 65 final, de 19 de febrero de 2020. Disponible en: https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_es.pdf Visitado el 1 de marzo de 2022.

Unión Internacional de Telecomunicaciones (2019), *United Nations Activities on Artificial Intelligence (AI)* disponible en internet en https://www.itu.int/dms_pub/itu-s/opb/gen/S-GEN-UNACT-2019-1-PDF-E.pdf Visitado el 1 de marzo de 2022.

— (2021), AI for Good, balance anual de resultados, disponible en https://www.aiforgood.itu.int Visitado el 1 de marzo de 2022.

— (2022) Focus Group on "Artificial Intelligence for Health", disponible en: https://www.itu.int/en/ITU-T/focusgroups/ai4h/Pages/default.aspx Visitado el 1 de marzo de 2022.

CAPÍTULO XVIII

INTELIGENCIA ARTIFICIAL Y EL FENÓMENO DE LA DESINFORMACIÓN: EL PAPEL DEL RGPD[1] Y LAS GARANTÍAS RECOGIDAS EN LA PROPUESTA DE LA LEY DE SERVICIOS DIGITALES

ANA GARRIGA DOMÍNGUEZ
Universidad de Vigo
agarriga@uvigo.es

1. INTRODUCCIÓN

En las últimas décadas se ha producido un gran desarrollo tecnológico, que tiene a la Inteligencia Artificial (IA) como protagonista. Su evolución, desde la IA simbólica, que se basaba en reglas predefinidas ejecutadas por una máquina, a la IA que se sirve de grandes conjuntos de datos (Big Data) y que se basa en el aprendizaje automático, ha tenido como consecuencia el incremento de su número de aplicaciones y de su capacidad para resolver problemas. Habida cuenta de la existencia de diferentes definiciones de IA, en este trabajo se adopta la recogida en la Recomendación sobre la Ética de la Inteligencia Artificial de la UNESCO de noviembre de 2021, según la cual, "los sistemas de IA son tecnologías de procesamiento de la información que integran modelos y algoritmos que producen una capacidad para aprender y realizar tareas cognitivas, dando lugar a resultados como la predicción y la adopción de decisiones en entornos materiales y virtuales".

La IA tienen múltiples aplicaciones y muchas de ellas emplean datos personales en alguna de las etapas de su ciclo de desarrollo y comercialización, ya sea en la fase de entrenamiento y de validación, ya sea en otras etapas posteriores de explotación. Así por ejemplo, en el ámbito de la salud, de la seguridad, del cálculo del riesgo financiero, en el ámbito de la publicidad personalizada, etc. No obstante, existen otros muchos sistemas de IA que no los utilizan en ninguna de sus fases de vida, ni en su desarrollo, ni en su explotación datos sobre personas, como pueden ser sistemas que se utilizan en el ámbito de la geología, la predicción de la climatología, etc. Naturalmente, en este trabajo, los sistemas de IA que nos interesan son los que afectan a las

[1] REGLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos).

personas y, especialmente, aquellos que son entrenados con datos personales o que los utilizan en alguna de las fases de su ciclo de vida.

Otra realidad que ha condicionado en los últimos años el desarrollo de los sistemas y herramientas del ámbito de la inteligencia artificial es la cantidad de información personal disponible, recopilada por los distintos servicios y aplicaciones de la sociedad de la información. Este hecho nos sitúa por sí mismo ante una nueva revolución tecnológica: el *Big Data.* Por otra parte, el desarrollo actual de la tecnología en el campo de la IA permite el perfilado ideológico individual, de la misma forma en la que pueden inferirse perfiles de otras clases, ya sea sobre preferencias de consumo, fiabilidad financiera, emocional o, incluso sobre orientación sexual (Sarigol, García y Schweitzer, 2014, 105). La elaboración de perfiles en las plataformas sociales a través de complejos algoritmos permite aplicar técnicas de micro-segmentación para elaborar información política personalizada. En este trabajo, se estudia la relación de estas técnicas con la difusión de noticias falsas y el fenómeno de la desinformación y su impacto en las libertades de expresión e ideológica y la institución de la opinión pública libre, que están en la base del sistema democrático. Se pretende aportar una reflexión sobre la obtención de perfiles, particularmente ideológicos, y su relación con las libertades de expresión e información, las *fake news* y el fenómeno de la desinformación a través de la red. Una de las ideas centrales de este trabajo es que existe una estrecha relación entre la garantía de la privacidad de las personas, especialmente a través del derecho fundamental a la protección de datos personales, y el ejercicio de las libertades de expresión e información y, consecuentemente, con garantía de la formación de una opinión pública libre que está en la base del sistema democrático.

2. INTELIGENCIA ARTIFICIAL Y PERFILADO IDEOLÓGICO: BIG DATA Y ALGORÍTMOS PREDICTIVOS

En la sociedad del control (Deleuze, 1992, 3-7) o sociedad de la transparencia (Byung-chul, 2013) el rastro que generamos a través de nuestra interacción con los diferentes servicios de la red es controlado y almacenado para diferentes fines. Esta interacción en el mundo virtual, pero también en el físico, es seguida y monitorizada por diversos vigilantes, que por diferentes razones e intereses, recogen y analizan nuestra actividad en los distintos servicios y redes de la Sociedad de la Información. En el mundo de la vigilancia líquida (Bauman/Lyon, 2013) cada uno de nuestros comentarios, acciones o intereses es susceptible de pasar a engrosar alguno de los muchos centros de datos que los Estados y las entidades privadas poseen y que en numerosas ocasiones constituyen su activo y objeto de negocio principal. A los sistemas basados en la utilización de *cookies* (Téllez Aguilera, 2001, 83 y ss.), que

posibilitan el funcionamiento de las denominadas *redes de seguimiento* a través de las cuales es posible seguir al usuario a medida que navega por la red, hay que sumar actualmente las tecnologías *fingerprinting*, que permite "una recopilación sistemática de información sobre un determinado dispositivo remoto con el objetivo de identificarlo, singularizarlo y, de esta forma, poder hacer un seguimiento de la actividad del usuario del mismo con el propósito de perfilarlo" (Aepd, 2019, 4). Nuestro rastro digital puede reunirse e interrelacionarse, con la consiguiente "transformación de datos en principio irrelevantes en un perfil peligrosamente público del ciudadano" (Drummond, 2004, 118).

Los avances en la minería y análisis de datos y el aumento masivo de la capacidad informática de procesamiento y almacenamiento, así como el número creciente de personas, dispositivos y sensores que están conectados por redes digitales ha revolucionado la capacidad de generar, comunicar, compartir y acceder a los datos (Tene/Polonetsky, 2012, 63). Cuando hacemos alusión a las tecnologías de *Big Data*, nos referimos, por un lado, a la gran cantidad de datos disponibles y, por otro, aludimos al conjunto de tecnologías cuyo objetivo es tratar grandes cantidades de información (Beltrán Pardo/Sevillano Jaén, 2013, 16 y ss.), empleando complejos algoritmos y estadística con la finalidad de hacer predicciones, extraer información oculta o correlaciones imprevistas y, en último término, favorecer la toma de decisiones. Para analizar estas inmensas cantidades de datos han surgido un conjunto de técnicas que hacen referencia a los sistemas de información y "que pertenecen al campo de la inteligencia artificial (y) recibe el nombre de «minería de datos»" (Ramos Bernal, 2012, 186) y que, utilizando la ingente cantidad de datos disponibles los analizan buscando patrones recurrentes y correlaciones.

El *Big Data* es alimentado por la propia interacción de los usuarios de los servicios de Internet y así, por ejemplo en las redes sociales, el usuario asume un doble papel, el de consumidor y el de creador y, de esta forma "son los propios usuarios los que crean una gran base de datos cualitativos y cuantitativos, propios y ajenos con información relativa a edad, sexo, localización e intereses" (Ortiz López, 2010, 24). Estas plataformas registran las acciones, estados de ánimo, interacciones y reacciones a los distintos estímulos a los que los usuarios son expuestos (Lanier, 2018). La combinación de estos datos, que el usuario aporta sobre el mismo y sobre terceros, permite la obtención de un perfil muy preciso de sus intereses y actividades y estos datos o el resultado de su procesamiento podrán ser utilizados con distintos fines,

sobre todo comerciales y de publicidad[2]. La publicidad comportamental online se basa "en conocer los hábitos de comportamiento del consumidor en la red con el propósito de ofrecerle publicidad personalizada" (Martínez Pastor, 2014, 291). La publicidad comportamental se basa en la observación continuada del comportamiento de los individuos para desarrollar un perfil específico, que permite proporcionar anuncios a medida de los intereses deducidos del comportamiento del usuario. A través del análisis de cada individuo, los anunciantes podrán dirigir a ese usuario "aquella publicidad que coincida con los gustos e intereses deducidos de dicho rastreo y análisis, incrementando así su eficacia" (CEPD, 2010). Internet, junto con las técnicas de publicidad basadas en el contexto o en los términos de búsqueda, "permite una verdadera personalización de la publicidad, mediante técnicas que tienen en cuenta información específica sobre el usuario concreto que está accediendo a un determinado sitio web" (Peguera Poch, 2010, 359).

Esta inmensa capacidad para buscar, agregar y realizar referencias cruzadas de los grandes conjuntos de datos (Boyd/Crawford, 2012, 662), que permiten extraer patrones de comportamiento y perfiles personales y que informan acerca de lo que somos y lo que hacemos (Craig/Ludloff, 2011, 6) hace posible una peligrosa y nueva filosofía de la anticipación, cuyo extremo sería el de las predicciones preventivas (Kerr/Earle, 2013). Una de las consecuencias de la publicidad comportamental o dirigida es que puede influir en los deseos de nuevas maneras, pero también puede mediar en los comportamientos reales de ciertos grupos sociales que, como los individuos, son alentados por retroalimentación para ajustarse a los patrones esperados (Lyon, 2014, 101). La libertad de elección y decisión de los individuos se verá directamente afectada, en la sociedad de consumo, sociedad sinóptica de adictos compradores/espectadores, "la obediencia al estándar (…) tiende a lograse por medio de la seducción, no de la coerción… y aparece bajo *el disfraz de la libre voluntad, en vez de revelarse como una fuerza externa"* (Bauman, 2013, 92). La focalización va a permitir transmitir mensajes específicamente diseñados para promover intereses económicos o comerciales, ideológicos o políticos, o de cualquier otra clase. Pues, el *targeting* tiene, como objetivo prioritario, orientar o dirigir al sujeto o a un grupo de personas en un sentido y con una finalidad determinados. Estos sistemas de IA, como la mayoría de los sistemas automatizados de toma, de ayuda a la decisión o los sistemas de recomendación, se basan cada vez más en la combinación de las tecnologías de Big Data y el aprendizaje automático y por ello resulta esencial el respeto a las

[2] Vid. Resolución sobre Protección de la privacidad en los servicios de redes sociales aprobada en Estrasburgo, los días 15 a 17 de octubre de 2008 en la 30 Conferencia Internacional de Autoridades de Protección de Datos y privacidad.

normas sobre protección de datos personales para proteger a las personas y garantizar sus derechos y libertades fundamentales. El Comité Europeo de protección de datos (CEPD) ha identificado una larga una serie de riesgos para los derechos y libertades de los usuarios de las plataformas y medios sociales, alguno de los cuales ya han sido mencionados. La elaboración de perfiles "relacionadas con la focalización podrían implicar una inferencia de intereses u otras características, que la persona no había revelado activamente, que socava la capacidad de dicha persona para ejercer el control sobre sus datos personales" (CEPD, 2021, 7). También, existe el riesgo de discriminación utilizando, o no, informaciones sensibles y, asimismo, como se ha venido insistiendo, existe un riesgo real de manipulación que podría afectar a cuestiones y procesos políticos, acentuando vulnerabilidades y emociones negativas y afectando a la autonomía, la libertad y a la salud psicológica, en especial en el caso de los menores. Pues, aprovechando determinados momentos en los que el análisis de la información revele estados emocionales determinados, se pueden "dirigir a la persona mensajes específicos y en momentos concretos a los que se espera que sea más receptiva e influir así subrepticiamente en su proceso de pensamiento, sus emociones y su comportamiento" (CEPD, 2021, 7). Finalmente, el CEPD refiere los riesgos para las libertades de expresión e información a través de los filtros burbuja, que limitan el acceso a las fuentes de información o tienen como resultado la autocensura. Ya que, "el uso de algoritmos para determinar qué información se muestra a qué personas puede afectar negativamente a la posibilidad de acceder a fuentes de información diversificadas en relación con un tema concreto" (CEPD, 2021, 8). Esta práctica tendrá consecuencias negativas para el pluralismo político y el debate público de ideas, pero también para el acceso a una información plural y diversa. Las herramientas de focalización "pueden utilizarse para aumentar la visibilidad de determinados mensajes y, al mismo tiempo, dar menos importancia a otros. El posible impacto adverso puede producirse a dos niveles. Por un lado, existen riesgos relacionados con los llamados «filtros burbuja», en los que la gente está expuesta a «más de la misma» información y encuentra menos opiniones, lo que provoca una mayor polarización política e ideológica" (CEPD, 2021, 8). Otro riesgo derivado del usos de estos mecanismos de focalización sería, como asimismo recoge el CEPD, el del exceso de información, que tendría como consecuencia que las personas al no poder conocer el grado de fiabilidad de la misma, no serían capaces de tomar una decisión con conocimiento de causa.

3. PLATAFORMAS SOCIALES Y MICRO-SEGMENTACIÓN *(MICROTARGETING)*: LOS RIESGOS DE LA FOCALIZACIÓN PARA LAS LIBERTADES DE EXPRESIÓN E INFORMACIÓN

Las plataformas sociales son servicios basados en la web que permiten a los individuos construir un perfil público o semipúblico dentro de un sistema delimitado. Ahora bien, el perfil o una cuenta en una red social "no es propiedad del usuario, es un espacio puesto a su disposición gratuitamente, a cambio de su disponibilidad a ser seccionado en partes comercialmente interesantes" (Ippolita, 2012, 64) y, como señala Bauman, en este ámbito, la socialización sigue las pautas del marketing y las herramientas electrónicas de la socialización digital "están hechas a la medida de las técnicas de marketing" (Bauman, 2007, 157). Para conseguir sus objetivos económicos las plataformas sociales diseñan su arquitectura y herramientas de formas muy concretas y este diseño fabrica nuestra realidad, la organiza y la orienta (Cardon, 2018, 21). Explica Lanier, que la información personal recolectada sirve para que los algoritmos establezcan correlaciones entre los datos de un mismo individuo y entre los de otras personas diferentes, constituyendo teorías sobre la naturaleza de cada persona midiendo y clasificando continuamente respecto de su predictibilidad y, posteriormente, "los algoritmos deciden lo que cada persona experimenta a través de sus dispositivos" (Lanier, 2018, 47).

La aplicación de la IA en estos ámbitos permite el perfilado ideológico individual y, a través de las técnicas de focalización podrá elaborarse información política personalizada, El desarrollo de los procesos de segmentación de mercados ha evolucionado hacia una segmentación psicográfica avanzada, que se basa en un algoritmo que determina una serie de rasgos demográficos y de actitud que permite distinguir a cada individuo para cada segmento objetivo y que permite hacer predicciones precisas de la reacción de la audiencia objetiva (Barbu, 2014, 44-45). De esta forma, la cantidad y calidad de la información personal que se encuentra en las redes sociales, permite a los anunciantes mejorar el alcance e impacto de su publicidad al dirigirse a grupos específicamente seleccionados y estructurados o, incluso, a individuos concretos para influir en su conducta (Barbu, 2014, 46). Obviamente, estas técnicas pueden utilizarse para vender un producto determinado, pero también para favorecer una determinada ideología y, asimismo, puede estar en la base del fenómeno de la desinformación a través de la red. El uso de perfiles permite determinar la información a la que vamos a tener acceso, limitando también nuestro derecho a recibir información veraz o siendo objeto de auténticas manipulaciones. A través de los distintos algoritmos utilizados por las plataformas sociales unos usuarios tienen acceso a un tipo de información y otros, en función de sus intereses o de su perfil ideológico, a otros contenidos diferentes (Bakshy/Messing/Adamic, 2015), aunque también pueden suponer

que las personas no se vean expuestas a informaciones que no encajen con su ideología. A través de los hilos de contenido personalizado, que se optimizan para captar a cada usuario, *"a menudo utilizando potentes estímulos emocionales que conducen a la adicción a las personas"* (Lanier, 2018, 48), se las puede manipular sin que sean conscientes de ello[3]. Se trata de modelos de negocio que basan sus ingresos en la venta de publicidad y por lo tanto necesitan captar la atención del usuario por lo que los algoritmos utilizados priorizarán aquellos contenidos que consigan este objetivo de forma más eficiente y el incremento de esta atención se consigue mejor a través de la amplificación de las emociones negativas frente a las positivas (Lanier, 2018, 42). Así, lo entendió también por el Parlamento británico en su informe, *"Disinformation and 'fake news': Final Report"*, del 14 de febrero de 2019, concluyendo que la proliferación de daños en Internet se hace más peligrosa al centrar mensajes específicos en los individuos como resultado de la «*micro-mensajería dirigida*», que a menudo se aprovecha y distorsiona la visión negativa de las personas sobre sí mismas y sobre los demás.

4. LA FUNCIÓN INSTRUMENTAL DEL DERECHO A LA PROTECCIÓN DE DATOS PERSONALES PARA GARANTIZAR LAS LIBERTADES DE EXPRESIÓN E IDEOLÓGICA

Existe una estrecha conexión entre la garantía de la privacidad de las personas, especialmente a través del derecho fundamental a la protección de datos personales, y libertades de expresión y a recibir una información veraz e ideológica. El escándalo Facebook-Cambridge Analytica evidenció una serie de prácticas que habrían afectado, al menos, a 50 millones de personas y, sobre cuyos datos personales almacenados por Facebook, se habrían elaborado perfiles individuales con fines de focalización política en la elecciones presidenciales de Estados Unidos de 2016 y en el referéndum sobre la permanencia en la Unión Europea del Reino Unido[4]. En la Resolución del Parlamento Europeo, de 25 de octubre de 2018, sobre la utilización de los datos de los usuarios de Facebook por parte de Cambridge Analytica y el impacto en la protección de los datos, se considera constatado que las fugas de datos de usuarios y el acceso concedido a aplicaciones de terceros sirvieron para utilizarse indebidamente en campañas electorales (Considerando A). Este caso,

[3] En ocasiones, bajo la coartada del experimento sociológico, se realiza sin tapujos una directa manipulación emocional de los usuarios. A lo largo de una semana durante el año 2012, Facebook experimentó con 689.000 usuarios sin su consentimiento para analizar su comportamiento alterando el algoritmo que selecciona las noticias que se ven de los amigos y, a través del tipo de noticias que mostraba a unos u a otros, positivas o negativas, para estudiar como influía en su estado de ánimo. (Kramer/Guillory 2014).

[4] De forma detallada se recogen en el Informe de la Cámara de los Comunes de 14 de febrero de 2019 "Disinformation and 'fake news': Final Report".

ilustra claramente "cómo una posible vulneración del derecho a la protección de los datos personales podría afectar a otros derechos fundamentales, como la libertad de expresión y la libertad de opinión y la posibilidad de pensar libremente sin manipulación" (CEPD, 2019, 1).

No es este un problema nuevo para las autoridades de protección de datos, que en el año 2005 adoptaron una Resolución sobre el uso de datos personales para la comunicación política poniendo de manifiesto la existencia de la realización invasiva de perfiles de personas, que las clasifica como simpatizantes, partidarios, adherentes o miembros de un partido, para intensificar la comunicación personalizada con grupos de ciudadanos. Para ello, las organizaciones políticas recopilarían una gran cantidad de datos personales que incluiría, además de los datos de contacto, informaciones sobre su actividad profesional y relaciones familiares, "datos sensibles relacionados con convicciones o actividades políticas o morales reales o supuestas, o con actividades de votación". Pero, si bien este no es un problema nuevo, lo cierto es que las posibilidades actuales de micro-segmentación y manipulación *online* basadas en las tecnologías de *Big Data* e inteligencia artificial que permiten la recolección, el almacenamiento, la combinación y el análisis de ingentes cantidades de datos personales hacen que el riesgo para los derechos de las personas sea hoy mucho más real y elevado. Como ha señalado el Supervisor Europeo de Protección de Datos existe una amenaza para los valores democráticos y los derechos fundamentales derivados de la incesante vigilancia a la que son sometidas las personas en el espacio digital por empresas y Estados y, esta disminución de su espacio íntimo tiene como consecuencia "un efecto alarmante sobre la capacidad y voluntad de las personas de expresarse y establecer relaciones con libertad, también en la esfera cívica, tan esencial para la salud de la democracia" (SEPD, 2018, 3).

El fenómeno de las noticias falsas es complejo, a las que contribuyen no solo las noticias falsas, sino también las cuentas falsas y *bots, "que amplifican la actividad e intensidad de los servicios"* (Lanier, 2018, 77) y es posible distinguirlas de un variado elenco de situaciones posibles que abarcarían desde las teorías de la conspiración hasta las informaciones erróneas pasando por las noticias de los medios de comunicación sesgados ideológicamente (Alcott/Gentzkow, 2016, 217 y ss.). Lo que caracteriza a las *«fake news»* es que éstas son intencional y verificablemente falsas (Alcott/Gentzkow, 2016, 213) y las empresas que las producen buscan el máximo beneficio a corto plazo para atraer el mayor número de «clics» y persiguen la viralización de la noticia y el aumento del trafico en la red *"porque eso es lo que impulsa la influencia y los ingresos por publicidad"* (Holiday, 2013, 290-291). Este fenómeno se ve potenciado porque su contenido puede ser retransmitido entre los usuarios sin necesidad de un filtrado de verificación de hechos o juicio editorial significativo por parte de

terceros, aprovechándose también del incentivo psicológico que supone el sesgo de confirmación (Alcott/Gentzkow, 2016, 211). Por la complejidad del fenómeno, la Comisión Europea prefiere centrar el debate sobre el problema de la desinformación, que incluiría la información inexacta, engañosa o falsa que ha sido diseñada y promovida para causar intencionalmente un daño (entre los que se puede incluir la amenaza a los procesos y valores democráticos, incluidas las elecciones) o con fines de lucro, que puede ser exacerbada por la forma en que el público y las comunidades reciben, se involucran y amplifican la desinformación (Comisión Europea, 2018, 10).

Las libertades de expresión y el derecho a la información veraz son, además de derechos fundamentales, garantías institucionales, es decir, instrumentos de los que se vale el sistema democrático para protegerse (Llamazares Calzadilla, 1999, 43 y ss.). La institución que garantizan es la opinión pública libre, fundamento del pluralismo político y elemento básico en un sistema democrático. Las libertades de expresión y de información cumplen por ello una función esencial de preservación del principio democrático y del pluralismo ideológico al permitir a los ciudadanos formar sus propias opiniones y convicciones, su conciencia individual y colectiva acerca de hechos y acontecimientos, así como participar en la discusión social sobre asuntos de interés público. En su dimensión objetiva, estas libertades actúan como elementos esenciales para establecer el necesario equilibrio entre poderes en las sociedades democráticas, contribuyen a realizar los fines del Estado porque son vehículos para la participación política y constituyen "un instrumento de control que tanto puede afectar al procedimiento de las tomas de decisiones como a la cualidad y legitimidad de las personas al frente de las instituciones políticas" (Soriano, 1990, 109). Por lo tanto, son "condición previa y necesaria para el ejercicio de otros derechos inherentes al funcionamiento de un sistema democrático (…). Para que el ciudadano pueda formar libremente sus opiniones y participar de modo responsable en los asuntos públicos, ha de ser también informado ampliamente de modo que pueda ponderar opiniones diversas e incluso contrapuestas" (STC 159/1986, de 16 de diciembre). Cuando la ciudadanía ejerce su derecho al sufragio elige sobre la base de un juicio que se construye sobre el conocimiento del que disponga de los asuntos públicos y su gestión. Y este conocimiento sobre asuntos de relevancia pública puede garantizar esa actuación libre de los ciudadanos pues, como nos recuerda el Tribunal Constitucional, "únicamente aquellas sociedades que pueden recibir informaciones veraces y opiniones diversas de cuanto constituyen los aspectos más importantes de la vida comunitaria, están en condiciones de ejercitar, después, sus derechos y cumplir sus deberes como ciudadanos, partiendo del principio esencial de que la soberanía nacional reside en el pueblo, del que emanan los poderes del Estado" (STC 173/1995, de 21 de noviembre).

La libre circulación de opiniones e informaciones se ve obstaculizada por bots y noticias falsas, pero también, cuando se aplican «burbujas de filtro», que a través de filtros invisibles, nos aísla sin percibirlo. Ya que, al desconocer la forma y los criterios según los cuales los servicios filtran la información que entra y sale, "es prácticamente imposible ver lo sesgada que es" (Parisier, 2017, 18). Como consecuencia de ello, cuando el entorno online se encuentra personalizado y micro-segmentado, los ciudadanos estamos expuestos a informaciones que refuerzan los sesgos ideológicos y es más difícil encontrar opiniones diferentes, lo que lleva "a una mayor polarización política e ideológica" (SEPD, 2018, 7). En este sentido, también el Parlamento Europeo[5] ha analizado los riesgos de la elaboración de perfiles utilizando macrodatos y, entre otras consideraciones, insta a la "Comisión y a los Estados miembros que velen por que las tecnologías basadas en los datos no limiten o discriminan el acceso a un entorno mediático pluralista sino que fomenten la libertad de los medios de información y el *pluralismo".* En su Informe de enero de 2018, el Grupo Consultivo sobre Ética del Supervisor Europeo de Protección de Datos señalaba, entre las amenazas para la autonomía individual, la difusión algorítmica o humana de noticias falsas que debilita la capacidad de los individuos para discriminar entre lo que es información fiable y lo que no lo es y, así también, los procesos democráticos estarían en riesgo de debilitarse a través de las prácticas de marketing político basadas en técnicas de micro-segmentación y elaboración de perfiles psicográficos; pues, las técnicas de micro-segmentación en el ámbito electoral cambia las reglas del discurso político, reduciendo el espacio para el debate y el intercambio de ideas.

Como consecuencia de estas prácticas, también la libertad ideológica podrá resultar afectada. Su papel es esencial en un Estado democrático y abarca "todas las opciones que suscita la vida personal y social (…) y para cuya efectiva realización es precisa la maduración intelectual en una mentalidad amplia y abierta" (ATC núm. 40/1999 de 22 febrero). La libertad ideológica se encuentra estrechamente vinculada a la dignidad humana puesto que "otorga dimensión moral a la vida humana en el sentido de que, en función de las creencias profesadas, el individuo puede orientar libremente el sentido de su existencia que adquiere así, en cuanto libremente determinada, dimensión moral" (Rollnert Liern, 2002, 66). Pero, además de fundamento de la autoderminación de la persona, la libertad ideológica es presupuesto del derecho de participación, del pluralismo político (Xiol Ríos, 2001, 19 y ss.) y requisito "requisito de funcionamiento del Estado democrático" Rollnert Liern,

[5] Resolución del Parlamento Europeo, de 14 de marzo de 2017, sobre las implicaciones de los macrodatos en los derechos fundamentales: privacidad, protección de datos, no discriminación, seguridad y aplicación de la ley.

2002, 70). En nuestro sistema de valores, la ideología es una cuestión privada e íntima. Se trata de un derecho complejo con distintas facetas o dimensiones y que en su dimensión externa, constituye "el reconocimiento de un ámbito de actuación constitucionalmente inmune a la coacción estatal" (STC 141/2000, de 29 de mayo) y en su dimensión interna, representa el "derecho a adoptar una determinada posición intelectual ante la vida y cuanto le concierne y a representar o enjuiciar la realidad según personales convicciones"(STC 120/1990 de 27 junio) y esta dimensión íntima o negativa de la libertad ideológica se concreta "por la determinación constitucional de que «nadie podrá ser obligado a declarar sobre su ideología, religión o creencias»" (STC 46/2001 de 15 febrero). Precisamente para garantizar este elemento negativo de la libertad ideológica y de conciencia, tanto el RGPD como la LOPDGDD[6] prohíben, como regla general aunque con determinadas excepciones, el tratamiento de los datos personales que revelen las opiniones políticas, las convicciones religiosas o filosóficas (artículos 9.1 de ambas normas). Ahora bien en el ámbito del *Big Data*, las posibilidades de elaboración de perfiles con el auxilio de la inteligencia artificial y el aprendizaje automático hace posible inferir las convicciones ideológicas y de conciencia de una persona sin que esta las haya hecho públicas, pudiendo "hallarse correlaciones que indiquen algo sobre la salud, las convicciones políticas, las creencias religiosas o la orientación sexual de las personas" (CEPD, 2018, 17)[7].

La relación entre todos estos derechos fundamentales está clara. Todos ellos contribuyen a garantizar la institución de la opinión pública libre y ello requiere que no se introduzcan interferencias y manipulaciones, se permita acceder a informaciones veraces y compartirlas e incluso discutirlas sin que los filtros automáticos y otros sistemas automatizados, basados en el perfil ideológico individual, limiten el alcance de las informaciones a los que la ciudadanía va a tener acceso. Por esta razón, existe una relación de interdependencia entre el derecho a la vida privada y las libertades de expresión e ideológica. Uno de los rasgos distintivos del derecho a la protección de datos personales es que cumple una importante función instrumental de garantía de otros derechos fundamentales. Así, se recoge en nuestro propio

[6] Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.

[7] El CEPD recoge el estudio de Kosinski, M., Stilwell, D. y Graepel, T.; *"Private traits and attributes are predictable from digital records of human behaviour"*, Proceedings of the National Academy of Sciences of the United States of America, volumen 110, nº 15, pp. 5802-5805. Dicho estudio, según se recoge por el Grupo del Artículo 29, *"combinó los «me gusta» de Facebook con información limitada procedente de encuestas y halló que los investigadores predijeron con exactitud la orientación sexual de un usuario varón en el 88% de los casos; el origen étnico de un usuario en el 95% de los casos; y si un usuario era cristiano o musulmán en el 82% de los casos"*.

texto constitucional (artículo 18.4 CE) y reiteradamente lo ha destacado el Tribunal Constitucional[8]. Desde el primer momento en que empezó a perfilarse su concepto, siempre se ha destacado cómo el derecho a la protección de datos personales cumple una misión relevante en atención a garantizar el ejercicio de otros derechos, constitucionales o no. Ello es así porque este derecho protege la libertad de elección de las personas, su derecho a no ser discriminadas y se encuentra directamente engarzada con la propia idea de dignidad y autorrealización humanas. De esta forma, el artículo 18.4 CE no solo «consagra un derecho fundamental autónomo a controlar el flujo de informaciones que conciernen a cada persona» sino también, se configura como "un derecho instrumental ordenado a la protección de otros derechos fundamentales" (STC 292/2000, de 30 de septiembre) entre los que se encontraría también el derecho fundamental a la libertad ideológica, tal y como reconoció expresamente el Tribunal Constitucional en su sentencia STC 76/2019, de 22 de mayo.

Que la protección a los derechos fundamentales frente al tratamiento de datos personales va más allá del propio derecho a la protección de datos o de la vida privada, es un hecho que se recoge y explica en el propio Preámbulo del RGPD. En su Considerando 4 se recogen otra serie de derechos que el Reglamento persigue proteger: el derecho al domicilio, el derecho de las comunicaciones, el derecho de la protección de los datos de carácter personal, el derecho de la libertad de pensamiento, de conciencia y de religión, el derecho de la libertad de expresión y de información, el derecho a la libertad de empresa, el derecho a la tutela judicial efectiva y a un juicio justo, y el derecho a la diversidad cultural, religiosa y lingüística, pues todos esos derechos están en juego. Esta referencia detallada a dichos derechos fundamentales en el preámbulo de la norma no es ornamental, sino que "anticipa el rigor con el que la Unión quiere precisar que sabe qué derechos están en juego y, por tanto, deben ser protegidos. Algo que reiterará una y otra vez a la largo de su articulado" (Fernández Hernández, 2018, 4) Para ello, se debe dotará a las personas de las herramientas que le permitan tener el control de sus propios datos.

[8] SSTC 11/1998, de 13 de enero (RTC 1998\11); 33/1998, de 11 de febrero (RTC 1998/33); 35/1998, de 11 de febrero (RTC 1998/35); 45/1998 de 24 de febrero (RTC 1998/45); 104/1998, de 18 de mayo; 198/1998 (RTC 1998/198), de 13 de octubre o 44/1999, 22 de marzo (RTC 1999/44).

5. LAS GARANTÍAS DEL RGPD: LA REGULACIÓN DE LA ELABORACIÓN DE PERFILES Y EL PAPEL ESENCIAL DEL PRINCIPIO DE TRANSPARENCIA

Los riesgos para los derechos de las personas para su libertad de elección, ante las posibilidades de discriminación y exclusión, o ante las consecuencias de predicciones erróneas o inexactas, se tratan de conjurar en el Reglamento General de Protección de Datos a través de la específica regulación de la elaboración de perfiles y el derecho a no ser objeto de decisiones basadas únicamente en tratamiento automatizados. La Recomendación (2010)13 sobre la protección de las personas con respecto al tratamiento automatizado de datos de carácter personal en el contexto de la creación de perfiles del Comité de Ministros del Consejo de Europa, ya había señalado los problemas que la técnica de creación de perfiles puede "suponer graves amenazas para los derechos y libertades de las personas" y, no solamente porque este se construye con los datos directamente proporcionados por el interesado, sino también por las posibilidades de generar nuevos datos personales por inferencia. Los efectos de estas operaciones serán especialmente graves cuando se realicen correlaciones utilizando datos sensibles o estados emocionales que permitirían un mayor grado de una manipulación.

En su artículo 22, el Reglamento garantiza el derecho del interesado a no ser objeto de una decisión basada únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzca efectos jurídicos en él o le afecte significativamente de modo similar. Se limitan las decisiones basadas únicamente en el tratamiento automatizado, es decir, aquellas que "representan la capacidad de tomar decisiones por medios tecnológicos sin la participación del ser humano" (CEPD, 2018a, 8).

En el propio RGPD, en el artículo 4.4, se indica que debe entenderse por «elaboración de perfiles»: toda forma de tratamiento automatizado de datos personales consistente en utilizar datos personales para evaluar determinados aspectos personales de una persona física, en particular para analizar o predecir aspectos relativos al rendimiento profesional, situación económica, salud, preferencias personales, intereses, fiabilidad, comportamiento, ubicación o movimientos de dicha persona física. En este precepto se recogen los tres elementos que determinan que estemos ante una elaboración de perfiles: que se trate de un tratamiento automatizado, que se evalúen aspectos personales de una persona física y que ese tratamiento automatizado se base en datos personales.

El artículo 22 se aplica tanto a la elaboración de perfiles como a la adopción de decisiones automatizadas, estén o no basadas en perfiles. Como ya se ha explicado, la elaboración de perfiles implica la recogida de información

sobre una persona o un conjunto de personas y la evaluación de sus características o patrones de comportamiento con el fin de clasificarlas, asignarlas a una determinada categoría o grupo con el objetivo de analizar o hacer predicciones sobre sus características, intereses, capacidades o comportamiento futuro. Sin embargo, como señala el Comité Europeo de Protección de Datos, "las decisiones automatizadas tienen un ámbito de aplicación distinto y pueden solaparse parcialmente con la elaboración de perfiles o derivarse de esta" (CEPD, 2018a, 8), ya que pueden llevarse a cabo con o sin elaboración de perfiles.

De acuerdo con la interpretación del CEPD, el artículo 22 del RGPD contiene una prohibición general de tomar decisiones individuales basadas únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzcan efectos jurídicos o efectos significativamente similares, si bien existen excepciones a esta norma general y, dichas excepciones, cuando se apliquen exigirán la adopción de medidas específicas para garantizar los derechos y libertades del interesado, así como sus intereses legítimos (CEPD, 2018a, 16). Como ya he mencionado, las decisiones automatizadas, incluida la elaboración de perfiles, pueden tener graves consecuencias para las personas y, en este sentido, la aplicabilidad del artículo 22 dependerá de que se produzcan o se puedan derivar esas consecuencias relevantes para la persona. Considera el CEPD que "los efectos del tratamiento deben ser suficientemente importantes como para ser dignos de atención" y considera que podría tener ese potencial aquellas decisiones que puedan "afectar significativamente a las circunstancias, al comportamiento o a las elecciones de las personas afectadas"; aquellas que tengan "un impacto prolongado o permanente en el interesado"; o en los casos más extremos, aquellas que puedan "provocar la exclusión o discriminación de personas" (CEPD, 2018a, 24). Con carácter general considera que la presentación de publicidad dirigida no tendrá ese efecto significativamente similar que exige el artículo 22. No obstante, en determinadas circunstancias sí podría tenerlo, en función de las circunstancias específicas de cada caso, atendiendo al "nivel de intrusismo del proceso de elaboración de perfiles, incluido el seguimiento de las personas en diferentes sitios web, dispositivos y servicios; las expectativas y deseos de las personas afectadas; la forma en que se presenta el anuncio; o el uso de conocimientos sobre las vulnerabilidades de los interesados" (CEPD, 2018a, 24).

El derecho recogido en el artículo 22 no es absoluto, estableciéndose una serie de excepciones. Este derecho no se aplicará cuando la decisión esté autorizada por el Derecho de la Unión o de los Estados miembros que se aplique al responsable del tratamiento, sea necesaria para la celebración o la ejecución de un contrato entre el interesado y un responsable del tratamiento o se base en el consentimiento explícito del interesado. En estos dos últimos

supuestos el art. 22.3 exige que el responsable del tratamiento establezca medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado.

Finalmente, en el art. 22 se establece, además, una limitación en razón de la naturaleza de los datos personales prohibiéndose la adopción de decisiones automatizadas basadas en datos sensibles o especialmente protegidos (origen étnico o racial, opiniones políticas, convicciones religiosas o filosóficas, afiliación sindical, datos genéticos, datos biométricos, datos relativos a la salud o datos relativos a la vida sexual o las orientación sexuales) salvo que el interesado haya prestado su consentimiento explícito y esta posibilidad no esté prohibida por el Derecho de la Unión o de los Estados miembros o cuando el tratamiento sea necesario por razones de un interés público esencial, sobre la base del Derecho de la Unión o de los Estados miembros, que debe ser proporcional al objetivo perseguido, respetar en lo esencial el derecho a la protección de datos y establecer medidas adecuadas y específicas para proteger los intereses y derechos fundamentales del interesado. En ambos casos deberán tomarse medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado.

Otro requisito para la licitud para este tipo de tratamientos es el cumplimiento de las obligaciones derivadas de los principios relativos al tratamiento del artículo 5 del RGPD. De entre ellos, un principio especialmente relevante es el de transparencia, ya que estos procesos suelen ser opacos para las personas o a éstas les cuesta comprender su funcionamiento, así como, la relevancia de sus aplicaciones y consecuencias. La obligación de que se facilite al interesado de forma sencilla, fácilmente accesible y en un lenguaje claro y fácilmente comprensible, toda la información relevante para él en el proceso de tratamiento de sus datos es especialmente pertinente en el RGPD ,en aquellas situaciones, "en las que la proliferación de agentes y la complejidad tecnológica de la práctica hagan que sea difícil para el interesado saber y comprender si se están recogiendo, por quién y con qué finalidad, datos personales que le conciernen, como es en el caso de la publicidad en línea" (Considerando 58).

El principio de transparencia está íntimamente ligado al derecho a recibir una información completa, clara y sencilla relativa a todos los aspectos esenciales de un tratamiento de datos personales y a las posibles consecuencias que se podrían derivar de ese tratamiento. Esta exigencia de transparencia se conecta con el establecimiento de un contenido pormenorizado del derecho de información y de las correlativas obligaciones informadoras del responsable del tratamiento. Así se recoge en el artículo 13 del RGPD, que obliga al responsable del tratamiento a adoptar las medidas oportunas para facilitar al interesado toda información relevante relativa al tratamiento de sus datos personales

incluida la existencia de decisiones automatizas y la elaboración de perfiles así como información significativa sobre la lógica aplicada, la importancia y las consecuencias previstas de dicho tratamiento para el interesado. Igualmente, el principio de transparencia obliga al responsable del tratamiento a garantizar que se le informa aún cuando los datos no se hayan obtenido directamente de interesado en los términos previstos en el artículo 14 y a garantizar el derecho de acceso, en el artículo 15, a la información relativa a la existencia de decisiones automatizadas, incluida la elaboración de perfiles y, al menos en tales casos, a la información significativa sobre la lógica aplicada, así como la importancia y las consecuencias previstas de dicho tratamiento.

Finalmente, es necesario destacar que la obligación de transparencia se configura como "una expresión del principio de lealtad en relación con el tratamiento de los datos personales plasmado en el artículo 8 de la Carta de Derechos Fundamentales de la Unión Europea" (CEPD, 2018b, 5). Así se expresa también en el Considerando 39 del RGPD que exige que cualquier tratamiento de datos personales deberá de ser lícito y leal de forma que al interesado le ha de quedar totalmente claro que se están recogiendo y utilizando sus datos y en la medida en que éstos son o serán tratados de forma que, como señala el Comité Europeo de Protección de Datos, las personas no se vean sorprendidas "en un momento posterior del uso que se ha dado a sus datos personales" (CEPD, 2018b, 8). El RGPD exige que la información que se facilite al interesado comprenda todos los aspectos de las operaciones de tratamiento incluyendo la relativa a los distintos agentes que intervengan como corresponsables y, asimismo, "la relativa a los fines compartidos o estrechamente vinculados, los períodos de conservación, la transmisión a terceros, etc." (CEPD, 2021, p. 30). Esto ocurrirá en los casos en los que tratamiento de datos de perfilado se haga por cuenta de una empresa tercera, que sería el cliente de la plataforma digital.

Las exigencias de transparencia se van a concretar igualmente en la garantía efectiva del derecho de acceso del interesado. Desde el punto de vista de quienes realizan el tratamiento de sus datos, este derecho se concretaría en la obligación de proporcionarle un sistema fácil y accesible para conocer entre otros aspectos: la identidad del focalizador y toda la información relativa a este tratamiento. Considera el Comité que, para cumplir los requisitos del artículo 15 del RGPD (derecho de acceso) y garantizar la plena transparencia, los responsables del tratamiento podrían "considerar la posibilidad de aplicar un mecanismo para que los interesados puedan comprobar su perfil, incluidos datos de la información y las fuentes utilizadas para elaborarlo"; pues, las personas deben poder "conocer la identidad del focalizador, y los responsables del tratamiento deben facilitar el acceso a la información relativa a la

focalización, incluidos los criterios de focalización utilizados, así como el resto de la información exigida por el artículo 15 del RGPD" (CEPD, 2021, 32).

Pero no solamente se debe aplicar el principio de transparencia en la elaboración de perfiles; es necesario también el máximo rigor en la aplicación de los demás principios relativos al tratamiento (minimización, limitación de la finalidad, exactitud y veracidad, etc.). También las obligaciones derivadas del principio de responsabilidad proactiva y, en virtud de este principio, el RGPD contiene otras garantías importantes respecto de los derechos de las personas. Así por ejemplo, la exigencia del artículo 35.2.a) de la realización de una evaluación de impacto relativa a la protección de datos cuando un tratamiento de datos personales suponga la "evaluación sistemática y exhaustiva de aspectos personales de personas físicas que se base en un tratamiento automatizado, como la elaboración de perfiles, y sobre cuya base se tomen decisiones que produzcan efectos jurídicos para las personas físicas o que les afecten significativamente de modo similar".

Finalmente, respecto de los datos que puedan afectar a la ideología (por ejemplo, obtenidos por inferencia), es necesario recordar que el artículo 9 del RGPD establece como regla general la prohibición del tratamiento de datos personales que revelen las opiniones políticas, salvo que el interesado haya dado su consentimiento explícito para uno o varios fines y siempre que el Derecho de la Unión o él estatal permitan levantar esta prohibición y el artículo 9.1 de la LOPDGDD, dispone que "a fin de evitar situaciones discriminatorias, el solo consentimiento del afectado no bastará para levantar la prohibición del tratamiento de datos cuya finalidad principal sea identificar su ideología, afiliación sindical, religión, orientación sexual, creencias u origen racial o étnico". No obstante, en el párrafo segundo de este precepto se añade que "lo dispuesto en el párrafo anterior no impedirá el tratamiento de dichos datos al amparo de lo dispuesto en el artículo 9.2 del RGPD. Para nuestro Tribunal Constitucional "las opiniones políticas son datos personales sensibles cuya necesidad de protección es, en esa medida, superior a la de otros datos personales. Una protección adecuada y específica frente a su tratamiento constituye, en suma, una exigencia constitucional" (STC 76/2019, de 22 de mayo).

El artículo 9 del RGPD establece que, para poder tratar este tipo de informaciones, será necesario (artículo 9. 2) haber obtenido el consentimiento explícito del interesado o que los datos hayan sido manifiestamente hechos públicos por el interesado. Y como señala el CEPD, en los procesos de focalización, además de estas condiciones habrá de contarse con una base jurídica adecuada según el artículo 6 y llevarse a cabo de conformidad con los principios fundamentales establecidos en el artículo 5 (CEPD, 2020, 35). Estos

requisitos son exigibles tanto si nos encontramos ante datos explícitos del interesado, como si se trata de datos inferidos o combinados, ya que, en la medida en que una plataforma social o un proveedor de medios sociales "utiliza los datos observados para clasificar a los usuarios como de determinadas creencias religiosas, filosóficas o políticas, independientemente de que la clasificación sea correcta o verdadera, sin duda deberá considerarse esta clasificación del usuario como un tratamiento de datos personales de categoría especial en este contexto" (CEPD, 2021, 37).

6. GARANTÍAS ESPECÍFICAS PREVISTAS EN LA PROPUESTA DE REGLAMENTO (UE) DE SERVICIOS DIGITALES[9]

Uno de los objetivos de la futura Ley de Servicios digitas es crear un entorno en línea seguro, predecible y confiable, y para que las personas puedan ejercer los derechos garantizados por la Carta de los Derechos Fundamentales de la Unión Europea, en particular, la libertad de expresión e información" (Considerando 3). Fija su atención en la publicidad en línea ya que "puede contribuir a generar riesgos significativos, desde anuncios publicitarios que sean en sí mismos contenidos ilícitos hasta contribuir a incentivar económicamente la publicación o amplificación de contenidos y actividades en línea que sean ilícitos o de otro modo nocivos" (Considerando 52). Por ello, se establecen medidas para garantizar a los usuarios un mayor control sobre el uso de sus datos personales y, además, se prohíbe la publicidad dirigida cuando se trate de datos sensibles y dirigida a menores. Para lograr sus objetivos, en la Propuesta se refuerza el principio de transparencia estableciendo obligaciones para los prestadores de los servicios de informar sobre cuándo y en nombre de quién se presenta la publicidad, sobre los principales parámetros utilizados para determinar que se les va a presentar publicidad específica, que ofrezca explicaciones reveladoras de la lógica utilizada con ese fin, también cuando se base en la elaboración de perfiles, que complementarán las exigencias del RGPD. Además se recogen obligaciones específicas para las plataformas en línea de muy gran tamaño (con más de 45 millones de usuarios) debido a su alcance, en particular expresado en el número de destinatarios del servicio, "para facilitar el debate público, (…) la difusión de información, opiniones e ideas y para influir en la forma en que los destinatarios obtienen y comunican información en línea" (Considerando 53). Por lo tanto, en la Propuesta, las nuevas obligaciones de transparencia permitirán a los usuarios estar mejor informados sobre cómo se les recomiendan los contenidos y elegir al menos una opción no basada en la elaboración de perfiles. También estará prohibido manipular las elecciones de los usuarios mediante los denominados patrones oscuros.

[9] Propuesta de Reglamento del Parlamento Europeo y del Consejo relativo a un mercado único de servicios digitales (Ley de servicios digitales) y por el que se modifica la Directiva2000/31/CE.

En el Capítulo III se regulan las obligaciones de diligencia debida para crear un entorno en línea transparente y seguro y se establecen las obligaciones adicionales para las plataformas en línea y las exigencias específicas para aquellas de muy gran tamaño. En los artículos 23 y siguientes se dispone la obligación de informar sobre el uso de medios automáticos con fines de moderación de contenidos; sobre la publicidad en línea que exige, entre otras, que de forma clara y en tiempo real se informe al usuario de que se trata de publicidad, de en nombre de quien se presenta y de los parámetros utilizados para seleccionarlo de forma específica.

Para las plataformas de muy gran tamaño se exige la evaluación de los riesgos sistémicos. Entre estos riegos sistémicos, habrán de considerarse necesariamente los riesgos de difusión de contenido ilícito a través de sus servicios o cualquier efecto negativo para el ejercicio de los derechos fundamentales a la vida privada y familiar, la libertad de expresión e información, la prohibición de la discriminación y los derechos del niño. En tercer lugar, deberán evaluarse los riesgos de manipulación deliberada de su servicio que pueda producir un "un efecto negativo real o previsible sobre la protección de la salud pública, los menores, el discurso cívico o efectos reales o previsibles relacionados con procesos electorales y con la seguridad pública" (artículo 26). Estos riesgos podrían derivarse, como se indica en el Considerando 58, de la creación de cuentas falsas, del uso de bots u otros comportamientos, total o parcialmente automatizados, que pueden dar lugar a la difusión rápida y extendida de información que sea un contenido ilícito o incompatible con las condiciones de una plataforma. Para reducir estos riesgos deberán actuar con la debida diligencia, adoptando las medidas necesarias y adecuadas, adaptando el funcionamiento o el diseño de sus sistemas de moderación, de sus sistemas algorítmicos de recomendación e interfaces en línea, entre otras medidas que permitan salvaguardar el orden público y los derechos de las personas.

A las plataformas en línea de muy gran tamaño se les impone también la obligación de permitir el acceso a sus datos (artículo 31) al coordinador de servicios digitales de establecimiento o a la Comisión, cuando lo soliciten de forma motivada y en un período razonable, especificado en la solicitud, acceso a los datos que sean necesarios para vigilar y evaluar el cumplimiento las obligaciones de la Propuesta de Reglamento, que incluirá el acceso a sus interfaces de programación de aplicaciones. Se establece así principio de responsabilidad algorítmica permitiendo a las autoridades europeas y nacionales el acceso a los algoritmos de las grandes plataformas en línea.

7. CONCLUSIONES

El fenómeno de la desinformación no es nuevo, las teorías de la conspiración, los rumores sin verificar, los bulos o las mentiras sobre hechos objetivos o sobre personas ya existían antes de Internet y del desarrollo de las plataformas sociales. Sin embargo, "la diferencia con las fake news de hoy es que, con las plataformas digitales que dan sostén a las «redes sociales» y permiten la masiva generación e intercambio de contenidos, la desinformación multiplica y disemina de forma exponencial en tiempo real sin espacio para la reflexión o corrección" (Gutiérrez, 2018). Por otra parte, los filtros burbuja que explotan el sesgo de confirmación limita nuestra visión del mundo y consecuentemente nuestra capacidad para conocer y comprender la realidad y los puntos de vista de aquellos que por disentir de nuestra opinión son relegados por un sistema de inteligencia artificial. Estos mismos algoritmos para maximizar su eficiencia refuerzan las emociones negativas que son más productivas para incrementar el tiempo que pasamos conectados y captar nuestra atención.

Los procesos y las herramientas tecnológicos que confluyen en este fenómeno tan complejo son de difícil compresión para el usuario medio y su funcionamiento es poco transparente. La propia lógica del diseño del modelo de negocio, que se basa en la recogida masiva de datos personales para ser reelaborados en el ámbito de la realización de perfiles predictivos, que buscan condicionar el comportamiento de los individuos con diversos fines publicitarios y de marketing, son mucho más eficientes si este es un proceso es opaco.

Los sucesivos escándalos relacionados con la influencia del fenómeno de la desinformación a través de las redes sociales han tenido como consecuencia que estos servicios adopten determinadas medidas correctoras para luchar contra las noticias falsas en sus medios y que los gobiernos incluyan en sus planes de ciberseguridad los riesgos de las campañas de desinformación. Así, por ejemplo, la Comisión Europea[10] , reconociendo "que las campañas masivas de desinformación en línea con motivos políticos, particularmente a cargo de terceros países, con el objetivo específico de desacreditar y deslegitimar las elecciones constituyen amenazas crecientes" para las democracias, adoptó una serie de medidas para garantizar las elecciones de 2019 al Parlamento Europeo, que exigían la aplicación rigurosa del RGPD y de la Directiva 2002/58/CE, que garantiza la confidencialidad de las comunicaciones electrónicas. También

[10] COMUNICACIÓN DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO, AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO Y AL COMITÉ DE LAS REGIONES. Garantizar unas elecciones europeas libres y justas de 12 de septiembre de 2018.

promovió la aprobación de un Código de buenas prácticas en materia de desinformación, entre cuyos firmantes del Código en octubre de 2018 estarían Google, Facebook, Twitter, Mozilla y las asociaciones empresariales que representan al sector de la publicidad.

Sin duda estas medidas son relativamente útiles para combatir la desinformación, pero, en mi opinión, se trata de soluciones insuficientes. Debe exigirse que los procesos de perfilado, que afecta a las oportunidades vitales de las personas, que dependen de la categoría en la que las hayan situado y que permite condicionar su conducta a través de la seducción, como señaló Bauman, sean completamente transparentes respecto de cómo se elaboran, para qué y para quienes, tal y como exige el RGPD. Cuando cualquier decisión se adopta con base en un perfil, incluida la información a la que tendremos acceso, el principio de transparencia es imprescindible para poder conocer quién diseña esas categorías en las que se nos clasifica, quién decide su significado y quién decide bajo qué circunstancias esas categorías serán decisivas (Lyon, 2014, 186). Pues, como ha indicado el Supervisor Europeo de Protección de Datos, la correcta aplicación de las normas europeas que garantizan la protección de los datos personales y la confidencialidad de las comunicaciones electrónicas "debería ayudar a minimizar los daños causados por los intentos de manipular a los grupos". No obstante, es imprescindible también la exigencia de responsabilidad a "los actores de ecosistema (digital) que se benefician de las conductas nocivas"[11], ya que la apelación a la transparencia no es suficiente.

Por ello, la aprobación del futuro Reglamento de Servicios Digitales es urgente. Sus previsiones, incrementando las exigencias de transparencia y prohibiendo determinados comportamientos online se presentan como instrumentos idóneos para luchar contra el fenómeno de la desinformación. El establecimiento de obligaciones específicas para las plataformas en línea y especialmente aquellas de muy gran tamaño así como la previsión de elevadas sanciones para las conductas más graves de hasta 6% de la renta o facturación anual del prestador de servicios intermediarios afectado (artículo 42) y la garantía del derecho a presentar una reclamación de los destinatarios del servicio (artículo 43) contribuirán a garantizar el respeto a los derechos fundamentales de las personas en su actividad en las plataformas sociales, en especial a su vida privada, a expresarse y a recibir una información veraz, contribuyendo al sostenimiento de los valores democráticos.

[11] Resumen del Dictamen del SEPD sobre la manipulación en línea y los datos personales. DOUE del 4 de julio de 2018.

8. BIBLIOGRAFÍA

Agencia Española de Protección de Datos (2019), *Estudio fingerprinting o Huella digital del dispositivo*, AEPD, Madrid.

Alcott, Hunt. y Gentzkow, Matthew (2017), "Social Media and Fake News in the 2016 Election", en: *Journal of Economic Perspectives*, 31, 2, 211-236.

Autoridades de Protección de Datos (2005), *Resolución sobre el Uso de Datos Personales para la Comunicación Política* de las, adoptada en la Conferencia de Montreaux del 14 al 16 de septiembre de 2005.

Bakshy, Eytan, Messing, Solomon y Adamic, Lada. A. (2015), "Exposure to ideologically diverse news and opinion on Facebook", en: *Science*, 348, 1130-1132.

Barbu, Oana (2014), "Advertising, Microtargeting and Social Media", en: *Procedia - Social and Behavioral Sciences*, 163, 44-49.

Bauman, Zygmunt (2007), *Vida de consumo*, Fondo de Cultura Económica, Madrid.

— (2013), *Modernidad líquida*, FCE, Buenos Aires.

Bauman, Zygmunt, David Lyon (2013), *Vigilancia líquida*, Paidós, Barcelona.

Beltrán Pardo, Marta y Sevillano Jaén, Fernando (2013), *Cloud computing, tecnología y negocio*, Paraninfo, Madrid.

Boyd, Danah. Kate Crawford (2012) "Critical questions for Big Data", en: *Information, Communication & Society*, 15, 1-18.

Byung-Chul Han (2013), *La sociedad de la transparencia*, Herder, Barcelona.

Cámara de los Comunes (2019), Informe *Disinformation and 'fake news': Final Report"*, de 14 de febrero de 2019.

Cardon, Dominique (2018), *Con qué sueñan los algoritmos. Nuestras vidas en el tiempo de los Big Data,* Dado ediciones, Madrid.

Comisión Europea (2018), *Final report of the High Level Expert Group on Fake News and Online Disinformation: "A multi-dimensional approach to disinformation*.

Comité Europeo de Protección de Datos (2010), *Dictamen 2/2010 sobre Publicidad comportamental* del Grupo de Trabajo del artículo 29, adoptado el 22 de junio de 2010.

— (2018a), *Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679*, adoptadas el 3 de octubre de 2017 y revisadas por última vez y adoptadas el 6 de febrero de 2018.

— (2018b), *Directrices sobre la transparencia en virtud del Reglamento (UE) 2016/679* del GT29, adoptadas el 29 de noviembre de 2017 y revisadas por última vez y adoptadas el 11 de abril de 2018.

— (2019), *Declaración 2/2019 sobre el uso de datos personales durante las campañas políticas*, adoptada el 13 de marzo de 2019.

— (2021), *Directrices 8/2020 sobre la focalización de los usuarios de medios sociales*, Versión 2.0, adoptadas el 13 de abril de 2021.

Craig, Terence, Mary E. Ludloff (2011), *Privacy and Big Data*, O'Reilly Media.

Deleuze, Gilles (1992), "Postscript on the Societies of Control", en: *October* 59, 3-7.

Drummond, Víctor (2004), *Internet, privacidad y datos personales*, Reus, Madrid.

Ethics Advisory Group (EAG) (2018): *Towards a digital ethics*, Report 2018.

Fernández Hernández, Carlos (2018), "El RGPD ¿última oportunidad para salvaguardar nuestros datos personales?", en: *Actualidad Civil* 5, 1-4.

Gutiérrez, Miren (2018), "Manual de fake news: El papel de los sesgos cognitivos", en: *eldiario.es*, 2 de diciembre de 2018.

Holiday, Ryan (2013), *Confía en mi, estoy mintiendo. Confesiones de un manipulador de los medios*, Empresa Activa, Barcelona.

Ippolita (2012), *En el acuario de Facebook. El irresistible ascenso del anarco-capitalismo*, Enclave de Libros, Madrid.

Kerr, Ian y Earle, Jessica (2013), "Prediction, preemption, presumption: how Big Data threatens big picture privacy", en: *Stanford Law Review Online* 65, 65-72.

Kramer, Adam, Guillory, Jamie E. y Hancock, Jeffrey T. (2014) "Experimental evidence of massive-scale emotional contagion through social networks", en: *Proceedings of the National Academy of Sciences of Unites States of America* 24, 8788-8789.

Lanier, Jaron (2018), *Diez razones para borrar tus redes sociales de inmediato*, Editorial Debate, Barcelona.

Llamazares Calzadilla, Mª. Cruz (1999), *Las libertades de expresión e información como garantía del pluralismo político*, Civitas-Universidad Carlos III, Madrid.

Lyon, David (2014), *Surveillance Studies. An overview*, Polity Press, Malden.

Martínez Pastor, Esther (2014), "La publicidad comportamental online y la protección de los datos personales", en: Valero Torrijos, Julián, *La protección de los datos personales en internet ante la innovación tecnológica*, Aranzadi, Cizur Menor.

Ortiz López, Paula (2010), "Redes sociales: funcionamiento y tratamiento de información personal", en: Rallo Lombarte, Artemi y Martínez Martínez, Ricard (coord.), *Derecho y redes sociales*, Civitas, Madrid, 23-36.

Parisier, Eli (2017), *El filtro burbuja: Cómo la web decide lo que leemos y lo que pensamos*, Taurus, Madrid.

Parlamento Europeo (2018), Resolución de 25 de octubre de 2018, *sobre la utilización de los datos de los usuarios de Facebook por parte de Cambridge Analytica y el impacto en la protección de los datos* (2018/2855(RSP)).

Peguera Poch, Miquel (2010), "Publicidad online basada en comportamiento y protección de la privacidad", en: Rallo Lombarte, Artemi y Martínez Martínez, Ricard (coord.), *Derecho y redes sociales*, Civitas, Madrid, 355-380.

Ramos Bernal, Antonio (2012), *Reflexiones sobre economía cuántica*, ECU (Editorial Club Universitario), Alicante.

Rollnert Liern, Göran (2002), *La libertad ideológica en la jurisprudencia del Tribunal Constitucional*, eCentro de Estudios Políticos y Constitucionales, Madrid.

Sarigol, Emre, García, David Y Schweitzer, Frank; "Online Privacy as a Collective Phenomenon", en: *Proceedings of the second edition of the ACM conference on Online social networks*, ACM, 95-106.

Soriano, Ramón (1190), *Las libertades públicas*, Tecnos, Madrid.

Supervisor Europeo de Protección de Datos (2018), *Opinion 3/2018, on online manipulation and personal data*, adoptada el 13 de marzo de 2018.

Téllez Aguilera, Abel (2001), *Nuevas tecnologías, intimidad y protección de datos*, Edisofer, Madrid.

Tene, Omer y Polonetsky, Jules (2012), "Privacy in the age of Big Data: a time for big decisions", en: *Stanford Law Review Online* 63, 63-69.

Xiol Ríos, Juan A. (2001), "La libertad ideológica o libertad de conciencia", en: AA.VV. *La libertad ideológica. Actas e las VI Jornadas de la Asociación de Letrados del Tribunal Constitucional*, Cuadernos y Debates nª 115, Centro de Estudios Políticos y Constitucionales, Madrid, 11-80.

CAPÍTULO XIX

DECISIONES AUTOMATIZADAS, DERECHO ADMINISTRATIVO Y ARGUMENTACIÓN JURÍDICA

LEONOR MORAL SORIANO
Universidad de Granada
lmoral@ugr.es

Los sistemas de decisión automatizada (Automated Decision-making, ADM) son tecnologías de Inteligencia Artificial diseñadas para asistir o incluso sustituir los juicios que hacemos los humanos. Aplicada a la esfera del Derecho, esta tecnología es utilizada por operadores jurídicos y singularmente por jueces. En estas circunstancias podría ser posible, a veces inquietante, un juez robot, imagen con la que se condensan nuestros temores sobre la irrupción de esta compleja tecnología en el razonamiento jurídico.

Mayor impacto, aunque solo sea por el número de ciudadanos afectados, tiene la aplicación de ADM en la actuación administrativa. En efecto, herramientas basadas en reglas, regresiones, analítica predictiva, *machine learning*, *deep learning*, o redes neuronales son algunas de las tecnologías que pueden conformar los sistemas de ADM y que la Administración Pública utiliza (bien en su forma de asistencia o de reemplazo) para decidir, es decir para adoptar actos jurídicos.

En esta contribución se argumentará que el Derecho Administrativo es un sistema normativo adecuado para el tratamiento de los sistemas de ADM cuando se utilizan en la adopción de actos administrativos. En concreto, los principios del Derecho Administrativo y normas relativas a la competencia, el procedimiento administrativo como garantía, y la motivación de los actos administrativos (esencial para su revisión) son algunos de las exigencias normativas de esta tecnología a las que debe responder cuando se utiliza en el razonamiento jurídico.

1. DE LA PROTECCIÓN DE DATOS AL DERECHO ADMINISTRATIVO

En un estudio empírico llevado a cabo por Ignacio Criado y publicado en *Eunomía*, este politólogo preguntó a los responsables de las políticas tecnológicas de cada departamento ministerial (Chief Information Officers o CIOs) cuáles eran los beneficios y las desventajas de la incorporación de la IA a las políticas públicas (Criado 2021). Con informes como los de la OCDE (2019) que conciben una Administración más efectiva y abierta, más transparente y participativa, que ofrece sus servicios 24/7, era de esperar que la valoración por

parte de la propia Administración fuera positiva. Y en general así fue, aunque no de forma tan entusiasta como cabría esperar[1].

Para impulsar la incorporación de la IA a las políticas públicas, el Gobierno español adoptó la Estrategia Nacional de Inteligencia Nacional (ENIA, en adelante)[2], donde se declara que la IA debe desarrollarse en sintonía con nuestras leyes y principios constitucionales (algo que afrancesadamente podría expresarse como "*va de soi*"). El marco jurídico español al que se refiere la ENIA no está desarrollado salvo en sus trazos más gruesos, a saber: su alineación con la normativa europea[3] y la protección de los derechos fundamentales, equidad en el acceso, así como prevención contra la discriminación.

Por lo que se refiere a la normativa europea, se quiere ir más allá de las guías éticas, fundamentales en el modelo de gobernanza estadounidense de la IA, pero cuya pobre eficacia Ulrich Beck compara con el freno de bicicleta en un avión intercontinental (Beck, 1988, 194 *apud* Haggendoff, 2020, 108). Además, en su propuesta jurídica, la UE está decidida a seguir la estela de la buena experiencia de la regulación de datos (RGPD) en la que Europa es un referente mundial[4].

Sin entrar ahora en el análisis exhaustivo del futuro marco jurídico europeo para la IA, basta indicar que se trata de un modelo regulatorio basado en la evaluación de riesgos con cuatro niveles identificados: riesgo mínimo, riesgo limitado, riesgo alto y riesgo inaceptable; además cuenta con un modelo de gobernanza institucional que replica el creado para la protección de datos. En este contexto, claramente los sistemas de ADM que utiliza la Administración

[1] Entre los beneficios que percibían los CIOs entrevistados, solo obtuvieron un aprobado (más de 5 sobre 10) aspectos como la eficiencia (6,7) y la digitalización (5,7), pero no superaron el examen otros elementos como la transparencia, la seguridad de datos, la participación ciudadana, ni tampoco la interoperabilidad. A la hora de valorar qué desventajas acarrea la incorporación de la IA a las políticas públicas, los responsables identificaron la opacidad algorítmica (5,3), los problemas éticos (5,1), la desconfianza (5) y el reemplazo humano (4,9). Estos indicios, de acuerdo con Criado (2021, 369), confirman que los responsables de las políticas tecnológicas en nuestra Administración Pública esperan resultados positivos en las operaciones de gestión pública (el enfoque de servicios) más que otras áreas de la actuación vinculadas con los sistemas de gobierno.

[2] La ENIA puede consultarse en el portal del Ministerio de Economía: https://portal.mineco.gob.es/RecursosArticulo/mineco/ministerio/ficheros/201202_ENIA_V1_0.pdf

[3] Se refiere al Reglamento del Parlamento y del Consejo por el que se establecen las normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión (COM(2021) 206 final, de 21 de abril de 2021). Recibidos los informes preceptivos, un año después de su publicación, la propuesta está siendo negociada en el Consejo en su primera lectura.

[4] Para conocer los distintos modelos de gobernanza en IA, permítaseme la remisión a Moral 2021.

Pública estarán sujetos a evaluación ya que dependiendo de su mayor o menor complejidad requerirán el análisis de datos o incluso la creación de perfiles. Por ejemplo, el chabot del 016 requiere datos (proporcionados por el usuario) para avanzar en el ofrecimiento de información o en la prestación del servicio demandado. De la misma manera, cuando solicitamos una beca para estudiar un Grado universitario, aceptamos que los datos tributarios se obtengan directamente de la AEAT. También el análisis masivo de datos y su cruce está detrás de actas de inspección cuando se detecta el impago de obligaciones a la Seguridad Social. En conclusión, las tecnologías de la IA, y en particular las de ADM, han sido analizadas principalmente desde el prisma de la protección de datos (Criado 2021, 357).

Este enfoque orientado a los datos está presente en las Directrices sobre decisiones individuales automatizadas que aprobó el Grupo de Trabajo sobre protección de datos del artículo 29 del RGPD[5] en tanto que la tecnología que utilizan se nutre de cualquier tipo de datos: aquellos ofrecidos por las personas afectadas, los observados acerca de sus personas, y los derivados o inferidos (un perfil de la persona).

Dicho enfoque también se vislumbre tras el modelo regulatorio propuesto por la Comisión en su Reglamento de IA. Los sistemas de ADM pueden considerarse como de alto riesgo de acuerdo con el Reglamento europeo porque pueden ser utilizados por las Administraciones públicas o entidades colaboradoras "para evaluar la admisibilidad de las personas físicas para acceder a prestaciones y servicios de asistencia pública, así como para conceder, reducir, retirar o recuperar dichas prestaciones y servicios" (Anexo III de la propuesta de Reglamento). Su utilización en el ámbito del ejercicio de la potestad sancionadora también debería ser considerado como de alto riesgo, si bien el tenor literal del Anexo III de la propuesta de Reglamento europeo sólo se refiere sistemas de IA utilizados en asuntos relacionados con ilícitos penales (no con infracciones administrativas)[6]. La propuesta de Reglamento establece los criterios que deben satisfacer los sistemas de IA de alto riesgo, entre ellos,

[5] Grupo de Trabajo sobre protección de datos del artículo 29 del RGPD, Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679, 17/ES, WP251rev.01.

[6] El Reglamento se refiere a sistemas de IA destinados a "determinar el riesgo de que se comentan infracciones penales"; "la fiabilidad de las pruebas durante la investigación o el enjuiciamiento de infracciones penales"; "predecir la frecuencia o reiteración de una infracción penal real o potencial con base en la elaboración de perfiles de personas físicas"; "elaboración de perfiles de personas físicas durante la detección, la investigación o el enjuiciamiento de infracciones penales"; "examinar grandes conjuntos de datos complejos vinculados y no vinculados, disponibles en diferentes fuentes o formatos, para detectar modelos desconocidos o descubrir relaciones ocultas en los datos".

como se ha indicado, los de ADM para el acceso a prestaciones y servicios. Los criterios hacen referencia a los datos y a la gobernanza de los datos (elemento central en el marco regulatorio); la documentación técnica; registros; transparencia y comunicación de información a los usuarios; vigilancia humana; y precisión, solidez y ciberseguridad (artículos 10 a 15).

A falta de que se apruebe la Ley europea de Inteligencia Artificial, y dado que los sistemas de ADM ya se utilizan en la actuación administrativa, son ciertas las palabras de Huergo, para quien "no existen normas que regulen el uso por las Administraciones de predicciones algorítmicas" (Huergo, 2021, 88); es decir, falta el enfoque del razonamiento jurídico. Ahora bien, la ausencia de un específico marco regulatorio de los sistemas de ADM no significa que no exista un sistema normativo adecuado que se debe aplicar a las decisiones automatizadas singulares, es decir, a aquellos actos administrativos en los que el operador jurídico ha utilizado (como asistente e incluso como reemplazo) un sistema de ADM. Este sistema normativo es el que facilita el Derecho Administrativo (Scassa, 2021), adoptando así una perspectiva en línea con el principio de sometimiento de la actividad administrativa a la Ley y al Derecho. En concreto, la atribución de competencia administrativa, el procedimiento administrativo como garantía, y la motivación del acto administrativo como requisito inescindible del derecho a la defensa, son algunos de los elementos normativos de las decisiones automatizadas, es decir, de aquellas decisiones en las que el operador jurídico ha utilizado sistemas de ADM.

2. LAS TECNOLOGÍAS DE LOS SISTEMAS DE ADM

Las tecnologías en las que se basan las decisiones automatizadas están estrechamente relacionadas con la extracción de datos y su tratamiento (correlación y cruce), con el cálculo de probabilidades, así como con la programación de procedimientos de decisión. Para ello se puede utilizar, en primer lugar y sin querer ser aquí exhaustiva, sistemas simbólicos donde se definen paso a paso el proceso de toma de decisión a partir de las reglas jurídicas y los hechos (datos) facilitados. Son árboles decisorios, más o menos complejos, que siguen la estructura "si-entonces". En segundo lugar, otra tecnología habitual de los sistemas de ADM es el *machine learning* mediante el entrenamiento (supervisado) de la máquina a partir de datos extraídos o facilitados. La tecnología sirve para analizar y cruzar datos masivos en una función que desempeña más eficientemente una máquina entrenada que un humano[7]. También puede utilizarse un aprendizaje de la máquina no

[7] Para Herbert Roitblat, los sistemas basados en *machine learning* pueden demostrar dificultades a la hora de identificar la relación entre causa y efectos (Roitblat, 2020, 344). Por ejemplo, Chen (2019) ha analizado millones de decisiones judiciales para extraer información sobre el impacto

(...)

supervisado, en cuyo caso, la finalidad es encontrar patrones para detectar fraudes en el cumplimiento de las obligaciones fiscales o a la Seguridad Social, por ejemplo.

La tecnología necesaria para los sistemas de ADM requiere, por lo tanto, datos y algoritmos. Para extraer datos en cantidad y calidad suficiente que sean relevantes para adoptar acto jurídico es necesario recordar tres aspectos recogidos en nuestro Derecho administrativo. En primer lugar, la interoperabilidad de los sistemas de información (artículo 3.2 LRJSP) para facilitar la interconexión y cruce de datos; en segundo lugar, la colaboración interadministrativa (artículo 3.1 LRJSP) de manera que todas las Administraciones Públicas compartan sus bases de datos; y en tercer lugar, el acceso a datos recabados por operadores privados, algo que ya existe a través de las obligaciones de información que la Administración Pública impone en sectores como las agencias de viajes, la comunicación, o las operaciones bancarias y financieras.

Por lo que se refiere a los algoritmos con los que se programan los sistemas de ADM que utilizan las Administraciones Públicas, nos encontramos con una dificultad fundamental desde el punto de vista semántico y conceptual (Hofmann 2021, 4): los sistemas de ADM están basados en un software diseñado por programadores informáticos quienes tendrán una concepción particular sobre el Derecho. Las normas jurídicas pueden ser inicialmente concebidas por no especialistas jurídicos como las líneas de programación; sin embargo, la interacción con otras normas jurídicas, principios, o precedentes no es un aspecto del Derecho que parezca fácilmente susceptible de computación.

En fin, en el procedimiento de toma de decisiones las Administraciones Públicas puede utilizar sistemas de ADM basados en distintas tecnologías, seleccionando una u otra según la fase del procedimiento[8]. En este sentido

de la fecha de nacimiento de los demandados, el nombre de las partes y del juez, el timbre de la voz (atractiva, masculina, ininteligible, etc.), el sexo, etc. Este tipo de investigación empírica está errada desde el inicio. Por ejemplo, supongamos que en un estudio nuestro objeto son las personas que tengan un número de DNI par y de ellas extraemos las siguientes conclusiones: hay más hombres con un DNI par que mujeres (o viceversa), son más los que están empleados que desempleados, hay más jóvenes que adultos, etc. Inferir que tener un número par de DNI es la causa de que se tenga más probabilidad de ser hombre, empleado, joven es simplemente absurdo.

[8] En un procedimiento administrativo el operador jurídico puede utilizar una combinación de sistemas de ADM. Si paulatinamente más fases del procedimiento están sujetas a ADM nos encontraremos ante una cyberdelegación, es decir, una forma de delegación del ejercicio de la potestad administrativa a favor del sistema automatizado. La cyberdelegación ha sido estudiada por Gogianes y Lehr (2017) y Cuéllar (2016) entre otros. Cuéllar advierte de que la dependencia de los programas informáticos, especialmente aquellos que se adaptan de forma autónoma (cajas negras), puede complicar aún más la deliberación pública sobre las decisiones administrativas,

(...)

Hofmann (2021, 4) indica que si bien los sistemas de ADM apenas se han utilizado en todas las fases de un procedimiento administrativo para la adopción de una decisión, hoy en día es más frecuente encontrarlos en las etapas iniciales de la actuación administrativa: en la planificación de la actuación y en la instrucción del expediente (como se tendrá ocasión de ver en el caso de las actas de inspección automatizadas de la Seguridad Social).

Ulrik Roehl (2022) ha identificado hasta seis tipos de uso de los sistemas de ADM en la actuación administrativa dependiendo del nivel de autonomía atribuido a la IA, es decir, del uso que el operador jurídico haga de la tecnología[9]. Esta clasificación funcional (no normativa) realmente identifica distintos niveles de interactuación entre el humano y el sistema de ADM utilizado, entre el operador jurídico y el algoritmo:

> Tipo A: automatización mínima. El operador jurídico decide sobre todos los aspectos de expediente administrativo y recibe la asistencia de tecnologías como un procesador de texto. Utilizará una *check-list*, instrucciones, y otro tipo de estándares decisorios que no están volcados en algoritmos.
>
> Tipo B: recuperación y tratamiento de datos. La decisión es compartida entre el operador jurídico y la tecnología. Ésta recaba, graba y presenta los datos relevantes para resolver el expediente. Por ejemplo, la concesión de becas al estudio requiere una tecnología que examine las solicitudes y extraiga los datos relevantes de las bases de datos de la Administración Pública.
>
> Tipo C: pasos procedimentales a seguir. Igualmente se produce una decisión compartida entre el operador y la tecnología. En este caso, la tecnología además de recuperar y seleccionar los datos relevantes, sugiere los siguientes pasos en el procedimiento. Por ejemplo, la tecnología utilizada en Estados Unidos para decidir las ayudas a los niños con discapacidad pertenece a esta categoría ya que el sistema evalúa las solicitudes: para los casos más sencillos se hace una recomendación automática de decisión, mientras que, para los casos más complejos, la tecnología sugiere que se evalúen directamente por el operador jurídico.
>
> Tipo D: decisiones asistidas. La decisión es compartida entre el operador jurídico y la tecnología. Ésta recaba, graba y presenta

porque serán pocos los observadores, si los hubiera, quienes serían completamente capaces de comprender cómo se llegó a una decisión determinada.

[9] Véase Roehl 2022 para una visión comprehensiva de las clasificaciones y tipologías de la automación elaborados por la doctrina.

algunos o todos los datos relevantes de un expediente y además sugiere un número limitado de soluciones o incluso una decisión específica. El ejemplo anterior sirve aquí también en tanto que la máquina propone o recomienda las decisiones posibles que puede adoptar el operador jurídico.

Tipo E: decisiones automatizadas. La tecnología, no el operador jurídico, es el autor principal de la decisión. Todos los aspectos se confían a la tecnología que opera automáticamente a partir de estadísticas y correlaciones, sin la asistencia del funcionario en el proceso de toma de decisión. Siguiendo el ejemplo de la concesión de becas, tras la recuperación y cruce de datos, el algoritmo decide la cuantía de la beca sin intervención del operador jurídico. Otro ejemplo lo ofrece la tecnología que identifica y notifica a los ciudadanos la deuda adquirida por haber recibido beneficios sociales indebidos; si el ciudadano no impugna la notificación en un determinado plazo, la tecnología comienza el procedimiento para la recuperación de la deuda. Algunos aspectos de estas decisiones automatizadas podrían incluso considerarse propias del siguiente tipo de tecnología.

Tipo F: Decisiones autónomas. De nuevo aquí el autor principal de la decisión es la tecnología. Todos los aspectos de la decisión administrativa se confían a la tecnología basadas en sistemas dinámicos de *machine learning* no supervisado, en los que el operador jurídico no interviene en el proceso de toma de decisión.

Cualquier decisión asistida por ADM es una decisión adoptada por un operador jurídico (Roehl, 2022, 49), incluso en el ámbito de las decisiones de los tipos E y F; ahora bien, en estos casos el humano no interviene en las fases iniciales para hacer una valoración del expediente, aunque sí en la última fase para revisar o incluso corregir la decisión a la que ha llegado la máquina.

Antes de avanzar en este capítulo conviene hacer una última precisión. Se refiere a la posible confusión entre decisiones basadas en ADM y la actuación administrativa automatizada del artículo 41 de la Ley 40/2015 de Régimen Jurídico del Sector Público. Nuestra LRJSP se refiere a una actividad que no requiere intervención directa del operador jurídico. Sin embargo, al explicar la tipología de decisiones automatizadas que adopta la Administración Pública, podemos concluir que toda actividad administrativa en la que haya una resolución administrativa exige la intervención humana, incluidas aquellas resoluciones basadas en sistemas ADM.

3. DERECHO ADMINISTRATIVO COMO EL SISTEMA NORMATIVO DE LAS DECISIONES BASADAS EN SISTEMAS ADM

Jennifer Raso (2021) ha propuesto en su contribución al libro colectivo *Artificial Intelligence and the Law in Canada*, utilizar el Derecho Administrativo como sistema normativo (sistema de normas y principios) para las decisiones administrativas basadas en ADM. Aunque se refiere al Derecho Administrativo canadiense, las diferencias entre sistemas de Derecho anglo-americano y de Derecho romano-germánicos no deben ser exageradas (Moral Soriano, 2008). A un lado y otro del Océano compartimos principios que conforman nuestro Derecho Público, y entre ellos, singularmente para nuestros fines, el principio del procedimiento administrativo como garantía del ciudadano.

En efecto, el procedimiento administrativo es un cauce ordenado de actuaciones que tiene como finalidad que la decisión sea conforme a Derecho, procedente y la más apropiada (si se trata de una decisión discrecional). Además, el procedimiento administrativo es, sobre todo, una garantía para que los interesados puedan defender así sus derechos e intereses legítimos.

Veamos algunos de los elementos del procedimiento administrativo que nos facilitan un sistema normativo idóneo para el tratamiento de las decisiones administrativas basadas en ADM.

3.1. Notificación y acceso al expediente

La Administración pública debe notificar al interesado el inicio del procedimiento administrativo, así como facilitarle el acceso al expediente (principio de contradicción y de transparencia). En el caso de una decisión administrativa basada en sistemas ADM, la notificación debería incluir información sobre el nivel de interrelación entre el sistema de ADM y el operador jurídico, el concreto sistema de ADM que se va a utilizar, si el sistema propone una decisión concreta, y cómo es su funcionamiento. Sin esta información, los interesados difícilmente podrán presentar alegaciones relevantes (en cualquier momento del procedimiento) o participar de manera significativa en el trámite de audiencia (Raso, 2021, 190).

En el Derecho Administrativo francés, la Administración tiene la obligación de notificar cuándo un acto administrativo ha sido adoptado utilizando un sistema de ADM. El art. L311-3-1 del Código de relaciones entre el público y la Administración (modificado por la *Loi pour une Rèpublique Numérique* de 2016, también denominada Ley Lemaire) establece que "*une*

décision individuelle prise sur le fondement d'un traitement algorithmique comporte une mention explicite en informant l'intéressé"[10].

Esta información (el que se esté utilizando un sistema de ADM) es primordial para el interesado desde el inicio del procedimiento, a fin de obtener información completa de cuáles son los elementos de juicio a los que acudirá el operador jurídico. Utilizando el ejemplo propuesto por Raso (2021, 184), imaginemos un ciudadano encarcelado por haber cometido un delito de agresión sexual. Para determinar si cumplirá condena en instalaciones de baja, media, o alta seguridad, los funcionarios de prisiones evalúan los riesgos utilizando tecnología que predice el comportamiento del preso a partir de su historial. Para ello, se utilizan datos aportados por el interesado y otros extraídos de bases de datos relativos a miles de personas cuyo riesgo ha sido evaluado anteriormente. El algoritmo da un valor que sugiere que el comportamiento futuro del preso presenta un riesgo medio para el resto de presos. Si el interesado supiera que el algoritmo da más relevancia a las correlaciones de datos que extrae de expedientes anteriores que de las respuestas que ha facilitado, podría conocer también los estándares en los que se basará la decisión; sin embargo, si desconoce el peso que se le dará a aspectos como el grupo étnico al que pertenece, el código postal, los antecedentes policiales, o las bases de datos que el sistema de ADM va a consultar, tendrá pocas posibilidades de oponerse y alegar evidencias significativas (Raso 2021, 190).

3.2. Audiencia

Presentar alegaciones en cualquier momento del procedimiento en su fase de instrucción, así como el trámite de audiencia presupone que exista intervención humana, que las alegaciones sean sopesadas por el responsable y, si son rechazadas que se consignen las razones para hacerlo. El artículo 22.3 del RGPD garantiza este derecho cuando establece que el responsable del tratamiento de los datos adoptará las medidas adecuadas para salvaguardar los derechos y libertades así como los intereses legítimos del interesado, como mínimo el derecho a obtener intervención humana por parte del responsable.

[10] Por su parte, el art. L300-2 califica el código fuente utilizado en las actuaciones administrativas como *documento administrativo*, y la doctrina de la *Commission d'Accès aux Documents Administratifs* (CADA) ha hecho lo propio con los algoritmos utilizados por la Administración pública, y la documentación técnica como puede ser el documento de especificación de requisitos de software. Entre otros, Gutiérrez David refiere el acceso al código fuente del programa de ADM utilizado por el Fondo Nacional de Subsidio Familiar para calcular las ayudas financieras de carácter familiar o sociales; tanto las cajas locales de asignación familiar, como los archivos SQL (*Structured Query Language*) del código fuente, y las especificaciones funcionales utilizadas para calcular las ayudas fueron consideradas por la CADA como documentos administrativos.

Ahora bien, la intervención humana (*human in-the-loop*) puede que no sea una garantía adecuada de los derechos e intereses legítimos. Raso vuelve a aludir en este punto a la decisión sobre el régimen de seguridad en una prisión (2021, 191). El operador jurídico, el funcionario de prisiones, tendrá que sopear la evaluación de las evidencias aportadas por el preso (inclinándose a un régimen de seguridad mínimo) y las aportadas por el sistema de ADM (que propone un nivel de seguridad media). Aunque la credibilidad del interesado no esté en entredicho, el sesgo de automación nos inclina a creer en los indicios generados por un algoritmo: los consideramos más objetivos y ajenos a los sesgos que acarreamos los humanos.

3.3. Motivación del acto administrativo

La motivación es un elemento esencial de los actos administrativos y está relacionada con el derecho fundamental a la tutela judicial efectiva; además, solo si el acto está suficientemente justificado podremos impugnarlo tanto en vía administrativa como contencioso-administrativa puesto que solo así conocemos las razones a rebatir; y si el acto no está suficientemente motivado, la impugnación procede por vulneración de un derecho fundamental.

Por otro lado, exigir de los actos administrativos que estén motivados es inherente a toda actividad jurídica de manera análoga a como exigimos de las decisiones judiciales que estén argumentadas. Los funcionarios públicos al igual que los jueces son operadores de un sistema normativo, el del Derecho, y deben fundamentar sus decisiones dentro de este sistema normativo, aportando no solo los hechos relevantes sino las razones apropiadas (leyes, reglamentos, precedentes, principios jurídicos, valores, etc.).

Puede ser que el término "motivación del acto administrativo" nos confunda sobre la naturaleza de la actividad que se lleva a cabo: la justificación jurídica. Quizás el desconcierto ha sido generado porque una razón no es lo mismo que un motivo, como nos recuerda García Figueroa, ya que los motivos etimológicamente hacen referencia a lo que nos *mueve*, mientras que las razones justifican la decisión. García Figueroa recurre al ejemplo de Otelo: "cabe decir que Otelo mató a Desdémona motivado (es decir, movido) por los celos, pero resultaría extraño decir que el moro de Venecia quitó la vida a Desdémona *justificado* por los celos" (2014, 142). Desenmascarada la ambigüedad razones/estímulos, motivar el acto administrativo es *justificarlo*, *argumentarlo*, *fundamentarlo* en razones pertenecientes a un sistema normativo, el del Derecho Administrativo.

Teniendo en cuenta que el órgano administrativo que motiva una decisión es un operador jurídico que razona en términos jurídicos, es esencial distinguir entre el principio de explicabilidad (recogido en la propuesta de

Reglamento de IA y presente en las guías éticas) y el requisito de la *justificación* jurídica. Recurriendo de nuevo al ejemplo de Raso (2021, 193), se debe *explicar* la tecnología del sistema ADM utilizado para conocer cómo se atribuye más peso o relevancia al perfil elaborado a partir de casos precedentes; ahora bien, se debe *justificar* (i.e., argumentar jurídicamente) por qué se utilizan sistemas de ADM en un concreto acto administrativo.

Explicabilidad y justificación no pueden confundirse porque pertenecen a ámbitos categoriales con direcciones de ajuste muy diferentes. El precursor de esta noción, de dirección de ajuste, es Tomás de Aquino, quien afirmó que la verdad es la correspondencia entre las cosas (*res*) y la mente (*intellectus*)[11].

La noción de direcciones de ajuste fue desarrollada por John Searle (1983, 7) quien distingue entre (i) ajustar nuestra mente al mundo, esto es, la dirección de la mente al mundo, y (ii) ajustar el mundo a nuestra mente, esto es, la dirección de mundo a mente:

> "La mejor ilustración que conozco de esta distinción la proporciona la señorita Anscombe. Supongamos que un hombre va al supermercado con la lista de la compra que le hizo su esposa en la que están escritas las palabras "judías, mantequilla, panceta y pan". Supongamos que mientras va con su carrito de compras seleccionando estos artículos, lo sigue un detective que anota todo lo que coje. Cuando salgan de la tienda, tanto el comprador como el detective tendrán listas idénticas. Pero la función de las dos listas será bastante diferente. En el caso de la lista del comprador, el propósito de la lista es, por así decirlo, lograr que el mundo coincida con las palabras; se supone que el hombre debe hacer que sus acciones encajen en la lista. En el caso del detective, el propósito de la lista es hacer que las palabras coincidan con el mundo; se supone que el hombre debe hacer que la lista se ajuste a las acciones del comprador. Esto se puede demostrar aún más

[11] Tomás de Aquino escribió: "La verdad consiste en la ecuación de [cosa y mente], como se dijo anteriormente. Ahora la mente, que es la causa de la cosa, se refiere a ella como su regla y medida; mientras que lo contrario es el caso de la cuenta que recibe su conocimiento de las cosas. Por tanto, cuando las cosas son la medida y el estado de la mente, la verdad consiste en la ecuación de la mente a la cosa, como sucede en nosotros mismos. Para según que una cosa es, o no es, nuestros pensamientos o nuestras palabras al respecto son verdaderas o falsas. Pero cuando la mente es la regla o medida de las cosas, la verdad consiste en la ecuación de la cosa a la mente; al igual que el trabajo de un artista se dice que es cierto, cuando se está de acuerdo con su arte. Ahora bien, como las obras de arte están relacionadas con el arte, por lo que son obras de justicia relacionados con la Ley con la que se otorguen. Por lo tanto, la justicia de Dios, que establece las cosas en el orden conforme a la regla de su sabiduría, que es la Ley de su justicia, se llama convenientemente verdad. Por lo tanto, también en los asuntos humanos hablamos de la verdad de la justicia". Cfr. Tomás de Aquino, *Summa Theologica*, Parte I, pregunta 21, el artículo 2.

> al observar el papel de un "error" en los dos casos. Si el detective llega a casa y de repente se da cuenta de que el hombre compró chuletas de cerdo en lugar de panceta, simplemente puede borrar la palabra "panceta" y escribir "chuletas de cerdo". Pero si el comprador llega a casa y su esposa le indica que compró chuletas de cerdo cuando debería haber comprado panceta, no podrá corregir el error borrando "panceta" de la lista y escribiendo "chuletas de cerdo" (Searle, 1979, 347).

La dirección de ajuste de la mente al mundo es la que siguen las ciencias como la informática: describe hechos. Nos dice qué es *normal*; nos explica cómo funciona el algoritmo o cómo está programado un sistema de ADM. La dirección de ajuste del mundo a la mente es la razón fundamental de los sistemas normativos como el Derecho: nos dice lo que debería ser. Nos dice qué es *normativo* y no debemos confundir normalidad con normatividad[12].

De hecho, bajo la dirección de ajuste de la mente al mundo, *describo* infracciones administrativas similares e incluso identifico contradicciones que conducen a declaraciones verdaderas / falsas; bajo la dirección de ajuste del mundo a la mente, *prescribo* la solución a las infracciones detectadas.

Estas dos direcciones de ajuste se hayan separadas en un plano conceptual (García Figueroa 2017, 113). Sin embargo, esto no significa que no haya lugar para la IA en el Derecho. Por el contrario, significa que el papel de la IA en el Derecho en general, y el de los sistemas de ADM y las decisiones administrativas en particular, deben ser analizado, teniendo en cuenta que son herramientas descriptivas en un ámbito prescriptivo. Son criaturas de otro mundo que también sirven al normativo.

De esta manera, los patrones de conducta y las previsiones que elabora tecnologías como el *machine learning* nos procuran premisas descriptivas. Esto establece un límite importante al papel que desempeñan en la argumentación jurídica. García Figueroa (2014, 86) dirige nuestra atención a la Ley de Hume[13]: es imposible derivar enunciados normativos exclusivamente de descriptivos; por tanto, es imposible derivar una decisión jurídica (juicio de deber) de solo una razón o premisa descriptiva. Esto significa que los sistemas de ADM son conceptualmente incapaces de justificar / argumentar / fundamentar decisiones jurídicas (actos administrativos o decisiones judiciales) por sí solos[14].

[12] Encontrar patrones nos descubre que es normal, y no propiamente normativo. Sobre la dicotomía normal / normativo (normalidad / normatividad) ver García-Pelayo (1968, 68).

[13] Uno de los mejores estudios relativos a la ley de Hume lo ha elaborado Celano (1994).

[14] Además, cuando el sistema de ADM está basado en sistemas de predicciones de resultados, presentan datos en términos de probabilidad de una manera que parece ser más neutral, más

(...)

En definitiva, el peligro no es la automatización de la actividad decisoria (que tiene sus ventajas en términos de eficiencia); el riesgo es que el realismo jurídico campe a sus anchas a rebufo del uso de la IA en el Derecho. Si lo relevante son las correlaciones que descubren las máquinas, los patrones imperceptibles para los humanos, las inferencias a partir del análisis de millones de datos, el Derecho será "papel mojado": "La actividad argumentativa de los juristas debería quedar sustituida por una actividad puramente ideológica, retórica o psicológica" (García Figueroa 2014, 99) encapsulada en patrones y probabilidades.

4. DOS CASOS DE ESTUDIO EN EL DERECHO ADMINISTRATIVO ESPAÑOL

4.1. Bosco

Canadá ha adoptado la Directiva sobre toma de decisión automatizada en el ámbito de la actuación administrativa que entró en vigor en enero de 2020. A pesar de contar con un texto normativo, indica Raso (2021, 187), se advierte una ausencia de doctrina jurisprudencial relativa al papel que juegan los algoritmos en las decisiones alcanzadas por la Administración Pública. La razón la podemos encontrar en la opacidad institucional: la Administración Pública no notifica que su decisión está basada en un sistema de ADM, y cuando lo hace, no explica su funcionamiento; y si lo explica, no todos tenemos los conocimientos o recursos para comprenderlo[15]. Por eso, cuando se impugna en vía judicial un acto administrativo basado en un sistema de ADM, las partes no aducen cuestiones relativas al papel de los algoritmos en el proceso de toma de decisiones (que desconocen), sino que esgrimen argumentos más asentados y conocidos como la defensa de los derechos humanos.

objetiva e incluso más precisa de lo que realmente es (Tashea, 2017). Por ejemplo, un funcionario de prisiones puede recibir un informe automatizado que indica que el acusado tiene un 80,2% de posibilidades de reincidir según el modelo de análisis jurídico (Surden, 2019, 1336). Sin embargo, según el modelo, 2 de cada 10 acusados no reincidirán. Por lo tanto, no es apropiado diferir una decisión legal sobre premisas descriptivas y engañosamente precisas (Surden, 2019, 1337) sin tener en cuenta los límites del modelo en términos de sesgo, discriminación y falta de transparencia.

[15] Monika Zalnieriute *et al.* (2021) identifica tres formas de opacidad: la primera es intencional y acontece cuando los sistemas de IA son tratados como bienes protegidos por derechos de autor, patentes o secretos comerciales, o bien cuando se utilizan datos sujetos a normas de privacidad o de protección de datos; la segunda forma de opacidad es el analfabetismo tecnológico ya que la mayoría de nosotros no seríamos capaces de extraer información útil del código base de la programación de sistemas de ADM; y la tercera forma de opacidad es realmente la consecuencia de las limitaciones humanas para entender y explicar cómo operan sistemas complejos, especialmente los de caja negra. Gutiérrés David indica que el argumento de la opacidad inherente es el que mayor calado está teniendo en la doctrina jurídica porque el acceso al código fuente es incomprensible para no expertos (2021, 177).

En España, un caso que incitará a elaborar doctrina jurisprudencial sobre los sistemas de ADM en la actividad administrativa es el asunto BOSCO. Estos son los hechos. La Fundación ciudadana CIVIO presta asistencia para la tramitación por parte de miles de consumidores del bono social, un pequeño descuento en la factura eléctrica. Las comercializadoras de energía consultan en la aplicación del Ministerio si sus clientes son beneficiarios del bono social, para lo que los solicitantes deben utilizar la herramienta elaborada al efecto. El sistema de ADM utilizado, denominado BOSCO, comprueba el cumplimiento de requisitos por los solicitantes y determina su elegibilidad, así como el tipo de descuento del que disfrutará el beneficiario. Los requisitos están relacionados con las vías de entrada al bono social: rentas bajas, familia numerosa, y beneficiarios con pensión mínima de incapacidad o de jubilación y que no cuenten con otros ingresos. De acuerdo con la taxonomía de los sistemas de ADM de Roehl, BOSCO pude ser considerada como una tecnología del tipo B (recuperación y tratamiento de datos) y D (decisión asistida).

No parece una herramienta de IA muy sofisticada, ni su programación parece que pudiera discrepar de la norma jurídica que establece los requisitos y las cuantías de las subvenciones. Sin embargo, CIVIO detectó casos en los que los solicitantes, cumpliendo los requisitos legales exigidos para disfrutar del bono social, el sistema ADM utilizado los rechazaba como beneficiarios. Uno de los problemas que detectó CIVIO fue la respuesta errónea a las viudas que solicitaban el bono social, porque no podían entrar por la vía de la pensión (aunque son pensionistas), y por lo tanto tenían que entrar por la vía del nivel de renta (si bien reciben una pensión, *no una renta*). Las solicitantes, CIVIO hizo la prueba, recibían dos respuestas a la solicitud: "no reúne los requisitos" (lógico, porque no tienen una pensión mínima de incapacidad o de jubilación) e "imposibilidad de comprobar los niveles de renta". Otro error fue detectado con las familias numerosas, cuyos miembros, independientemente de su nivel de renta, tienen derecho al bono social, y por lo tanto, no tienen que permitir el acceso a los datos de sus ingresos puesto que la renta no es un requisito para ser beneficiario. Aún así, si alguno de los miembros de la unidad familiar no marca la casilla que permite acceder a los datos de renta, BOSCO rechazaba su solicitud por "imposibilidad de cálculo"[16].

La aplicación de sistemas de ADM, incluso si se trata del tipo B de la taxonomía de Roehl (2022), esto es, sistemas de recuperación y tratamiento de datos, pueden conducir a decisiones administrativas contrarias a lo establecido por la normativa, lo que nos pone en guardia de que, incluso en el caso actos

[16] Como indica Gutiérrez David (2021, 164), la idea que subyace en el acceso al código fuente de BOSCO es comprobar si dicho programa está correctamente diseñado y si sus parámetros funcionales cumplen o no con las finalidades prevista en la norma jurídica que ejecuta BOSCO.

administrativos reglados, el acceso al algoritmo o al árbol de decisión es esencial para conocer la legalidad de la actividad administrativa. Siguiendo a Huergo (2021, 89), cuando el acto es reglado la utilización de sistemas de IA facilitan la automatización de la actividad administrativa y con ello su eficiencia. Su cometido es comprobar de manera automática el cumplimiento de requisitos que están establecidos por las normas jurídicas que regulan, por ejemplo, beneficios sociales: niveles de renta de la unidad familiar, expediente académico, número de créditos universitarios matriculados, no tener pendiente deudas con la Agencia Tributaria, etc. Considera Huergo que en estos casos el sistema de IA utilizado no determina el contenido de la actuación administrativa (Huergo 2021, 90), precisamente porque no hay margen de maniobra para la Administración pública. Pero los errores detectados en el funcionamiento de BOSCO nos convencen de que incluso en el caso de decisiones regladas los sistemas de ADM deben estar sometidas a control.

Ante la situación detectada por CIVIO, en septiembre de 2018 esta Fundación solicitó a través del Portal de Transparencia al amparo de la LTAIBG, la especificación técnica de BOSCO, el resultado de las pruebas realizadas para comprobar si cumplía con la especificación funcional, el código fuente de la aplicación y cualquier otro elemento que permitiera conocer el funcionamiento de la aplicación. Ante el silencio de la Administración, la Fundación interpuso una reclamación ante el Consejo de Transparencia y Buen Gobierno (CTBG). En su tramitación, la Administración rechazó facilitar la información requerida alegando motivos de seguridad pública, secreto profesional y propiedad intelectual e industrial, así como protección de datos personales, y riesgo de sufrir ataques informáticos si se conocieran las vulnerabilidades del programa[17]. El CTBG en su resolución 701/2018 de 18 de febrero de 2019, estimó parcialmente la reclamación de CIVIO e instó al Ministerio para la Transición Ecológica a que remitiera la información relativa a la especificación técnica, los resultados de las pruebas, y cualquier otro susceptible de ser entregado. No estimó, sin embargo, el acceso al código fuente[18].

Mediante el análisis de la información funcional, en concreto del árbol de decisión, CIVIO ya detectó errores en el procedimiento de decisión, pero necesitaba acceso al código fuente, por lo que interpuso un recurso contencioso-administrativo contra la resolución del CTBG. En el escrito de interposición de la demanda, CIVIO ejemplifica la necesidad de acceder al código fuente con una imagen efectista: si el código fuente consiste en una suma

[17] Son manifestaciones de la opacidad intencional de la que habla Zalnieriute *et al*. (2021).

[18] Sobre el acceso al código fuente que utiliza la Administración Pública y un extenso análisis al caso CIVIO se puede consultarse Gutiérrez David (2021).

simple 10+20+30+40+50, el resultado será 150. Si la Administración Pública entrega 150 y dice que es el resultado de una suma, la afirmación es correcta, pero no se conocerá el número de sumandos (qué y cómo se han valorado los diferentes factores que entran en juego).

Además, al no tener acceso al código fuente, se priva a los ciudadanos de conocer la motivación de la resolución. Pero la motivación (o indefensión) no fue el argumento principal al que apeló CIVIO. Más bien sostuvo que el código fuente es ley[19], y alegó vulneración del principio de legalidad (art. 9.3 CE), ya que, al programar, el ingeniero está reescribiendo la norma jurídica, a veces con clamorosos errores como en el caso de BOSCO[20].

El Juzgado Central de lo Contencioso-Administrativo desestimó la impugnación basada en el derecho de acceso al código fuente. Reflexiona, brevemente, sobre el papel de los sistemas de ADM en el procedimiento administrativo e indica que:

> "Aplicando al presente asunto los preceptos inmediatamente transcritos, debemos de considerar que el Ministerio para la Transición Ecológica, al reconocer el derecho al bono social, ajusta su actuación a dicha normativa, dictando el correspondiente acto administrativo. Y para ello utiliza una aplicación informática, denominada "sistema de información BOSCO", que se inserta en una

[19] Esta posición ya se ha sostenido en Derecho italiano, por el Consejo de Estado (similar a nuestra Sala 3 del Tribunal Supremo) en sentencia de 8 de abril de 2019 quien ha establecido que el algoritmo es una regla jurídica general sujeta a los mismos principios de transparencia y accesibilidad que se aplican a las normas. El asunto surge en un procedimiento de provisión de puestos de trabajo de personal docente, cuando los recurrentes detectan que la decisión adoptada no respondía ni al nivel educativo ni a la zona geográfica solicitada. La asignación de los puestos había sido determinada por un algoritmo cuyo funcionamiento no se hizo público. Además, las resoluciones administrativas carecían de motivación. Considera, como ya se ha adelantado, que el algoritmo es una regla jurídica general, y que como tal:

a) La regla, aun declinada de forma matemática, posee pleno valor jurídico y administrativo. En ese sentido, está sometida a los principios generales de la actividad administrativa, como aquellos de publicidad, transparencia, razonabilidad y proporcionalidad.

b) No cabe dejar espacio a la discrecionalidad en la aplicación del algoritmo.

c) La Administración ha de velar por los intereses en presencia, realizando pruebas, actualizaciones y sistemas de perfeccionamiento del algoritmo, en especial en el caso de aprendizaje progresivo y de *deep learning*.

d) Desde el punto de vista del control judicial, el algoritmo o software se entiende a todos los efectos como un acto administrativo informático y el órgano jurisdiccional ha de evaluar la corrección del proceso automatizado en todas sus vertientes.

[20] De manera similar, Boix Palop (2020) también defiende que los algoritmos utilizados por la Administración Pública en el proceso de toma de decisiones son regulaciones jurídicas porque cumplen las funciones propias de una norma jurídica.

> fase del procedimiento administrativo, cuyo objeto es verificar el cumplimiento de los requisitos establecidos previamente por la normativa citada.
>
> Siendo lo anterior así, no puede considerarse que el acto administrativo se dicte por una aplicación informática, sino por un órgano administrativo, y en caso de que el destinatario de dicho acto esté disconforme con el mismo, podrá impugnarlo en vía administrativa, y en vía judicial.
>
> Por tanto, la legalidad del acto administrativo no está justificada por la aplicación informática que instrumentalmente se utiliza en una fase del correspondiente procedimiento administrativo, sino por la normativa que regula la materia" (FJ 3).

Por lo tanto, el acto administrativo es dictado por un órgano, no por la aplicación. Además, si el interesado no está conforme puede impugnarlo (en vía administrativa y judicial). Ahora bien, si no se tiene acceso al sistema de ADM, y por lo tanto a todos los motivos o razones por las que se adopta dicho acto, difícilmente podrá ser impugnado. Por ejemplo, en el caso del sistema BOSCO, conocer su programación parece clave para determinar si el sistema de ADM se separa de (i.e., reescribe) la norma jurídica, si tiene en cuenta elementos de juicio distintos a los considerados en la norma (cruza datos con otras bases de tatos que en principio no están previstas en la regulación aplicable), o si pondera distintos aspectos de un caso y cuáles son los criterios utilizados. En fin, la distinción entre algoritmos predictivos y algoritmos deterministas o de aprendizaje automatizado no nos debiera inducir al error de suponer que en el primer caso la decisión ha sido elaborada con una herramienta informática que no tiene influencia sobre el contenido de acto administrativo. BOSCO ilustra la realidad de que incluso en el caso de sistemas de ADM basados en algoritmos predictivos, pueden existir casos difíciles o problemáticos en los que la aplicación rigurosa de la norma arroja un resultado diferente de la máquina.

4.2. Potestad sancionadora automatizada

La Administración Pública está incorporando en su actuación sistemas de ADM que encajan en la descripción de decisiones automatizadas y decisiones autónomas que propone Roehl (2022); es el caso de la actuación en el ámbito de las infracciones y las sanciones en materia social y seguridad social reformada por Real Decreto 688/2021, de 3 de agosto, por él se modifica el Reglamento general sobre procedimientos para la imposición de sanciones por infracciones de orden social y para los expedientes liquidatarios de cuotas de las Seguridad Social (RGPSOS en adelante).

Esta última modificación operada en el RGPSOS incluye la inspección automatizada (arts. 43 y ss.) como resultado de la incorporación de un sistema de ADM que facilita el análisis y cruce de datos provenientes de distintas Administraciones públicas como la Agencia Tributaria, el Servicio Público de Empleo, y el Instituto Nacional de la Seguridad Social, entre otros, para la detección de incumplimientos legales. El sistema de ADM está basado en un algoritmo predictivo: detecta indicios de infracciones administrativas sobre los que se decide recabar y cruzar datos para evacuar la correspondiente acta de infracción o incluso iniciar un expediente sancionador. El sistema utilizado crea perfiles del potencial defraudador a quien la máquina decide inspeccionar[21].

El procedimiento promovido por actuación automatizada se iniciará por orden del Director del Organismo Estatal de ITSS, que se emitirá para la realización de cada conjunto de actuaciones indicando los criterios a seguir en la preparación y ejecución de la actuación, así como el órgano encargado de su realización (art. 44.1 RGPSOS).

Las actas de infracción automatizadas deben reflejar los hechos comprobados de forma automática por el cruce de datos, con expresión de aquellos que sean relevantes a efectos de la tipificación de la infracción, los medios utilizados para la comprobación de los hechos que fundamentan el acta y la indicación expresa de que se trata de una actuación administrativa automatizada iniciada mediante expediente administrativo (art. 45 RGPSOS)[22]. Desde el punto de vista sustantivo, nada dice la norma sobre el ámbito en el que se aplicarán los algoritmos, más allá de las reclamaciones de deuda que ya están automatizadas como nos recuerda Goerlich Peset, por lo que parece plausible

[21] Un precedente del uso de sistemas de ADM en el ámbito de la lucha contra el fraude lo ofrece el Derecho holandés en el caso SYRI (*Systeem Risicoindicatie*); se trata de una herramienta de ADM que detecta fraudes en la percepción de subsidios y beneficios sociales. Hay una sentencia de 5 de febrero de 2020 del Tribunal de Distrito de la Haya donde concluye que se había vulnerado el artículo 8 del Convenio Europeo de Derechos (derecho al respeto de la vida privada y familiar) por parte de las autoridades holandesas. En concreto, el pronunciamiento venía referido al uso de un instrumento para detectar fraude, instrumento de naturaleza algorítmica que realizaba predicciones sobre la base de datos insertados en el sistema. El órgano jurisdiccional holandés consideró que esta actuación administrativa contravenía el artículo 8.2 del Convenio, una vez realizada la correspondiente ponderación entre los derechos e intereses concurrentes, por cuanto el algoritmo no era transparente ni verificable.

[22] Parece que el RGPSOS no ha atribuido presunción de certeza a las actas automatizadas, cosa que expresamente reconoce en el caso las actas de los funcionarios de la Inspección de Trabajo y Seguridad Social (art. 15 RGPSOS); sin embargo, Goerlich Peset (2021) advierte que esta eficacia probatoria privilegiada puede "entrar por la puerta de atrás" cuando, a tenor de las alegaciones que haga el interesado, tenga lugar la intervención de un funcionario que evacuará informe (art. 47 RGPSOS).

que las actas automatizadas puedan ser utilizadas en otros ámbitos del procedimiento sancionador en el orden social (Goerlich Peset 2021, 29).

El acta de infracción automatizada se notificará al presunto responsable. El art. 47 RGPSOS describe tres escenarios.

> Primero, si en el plazo para formular alegaciones el sujeto procede al pago de la sanción propuesta, se dará por concluido el procedimiento, lo que lleva implícito el reconocimiento de responsabilidad y la renuncia al ejercicio de cualquier acción, alegación o recurso en vía administrativa.
>
> Segundo, si el sujeto no formula alegaciones y no procede al pago de la sanción, el acta de infracción automatizada será considerada propuesta de resolución sancionadora.
>
> Tercero, si el sujeto formula alegaciones, se debe designar a un actuante con funciones inspectoras para que informe sobre las mismas. Tras la emisión del informe ampliatorio se continuará con la instrucción del procedimiento sancionador.

Como este uso de las actas de infracción automatizadas ya opera en ámbitos como las infracciones de tráfico, quizás pase desapercibida la extraordinaria innovación que opera (Goerlich Peset 2021): el sistema de ADM elige objetivos y detecta el fraude a partir de perfiles que elabora. Más aún, no se podrá impugnar la creación del perfil, porque en la motivación del acto automatizado no está previsto facilitar el algoritmo. Nos encontramos así ante una actividad administrativa opaca de imposible verificación por parte del interesado (el juez no está mejor preparado), soslayando con ello los derechos de contradicción y defensa.

Paradójicamente, a ello le tenemos que añadir la opacidad inherente (Gutiérrez David 2021, 177). El concepto de *caja negra* en el contexto de la actividad administrativa asistida por sistemas de ADM se aplica por extensión "no sólo para los algoritmos de *machine learning* o *deep learning*, sino a cualquier modelo total o parcialmente automatizado de toma de decisiones, al margen del tipo de algoritmo implementado, cuando no es posible verificar la corrección y la adecuación a Derecho de las decisiones adoptadas por el modelo" (Gutiérrez David, 2021, 166). En estos casos, el código es incomprensible para los no expertos, y el interesado no estará en mejor posición para defender sus derechos si la motivación del acto administrativo incluye el acceso al código.

El Derecho administrativo francés sí ha vinculado la motivación del acto administrativo con la mención explícita a la finalidad perseguida por el tratamiento algorítmico (art. L311-3-1 del Código de relaciones entre el público

y la Administración). Esta mención explícita se concreta en el desarrollo reglamentario[23] (art. R311-3-1-1): "derecho a obtener la comunicación de las reglas que definen el comportamiento y sus principales características de aplicación, así como las modalidades de ejercicio de este derecho a la comunicación y de revisión, si corresponde, ante la Comisión de Acceso a los Documentos Administrativos". En todo caso, la información deberá incluir de forma inteligible "(1) el grado y el modo de contribución del tratamiento algorítmico a la toma de decisión; (2) los datos tratados y sus fuentes; (3) los parámetros de tratamiento, y si procede, su ponderación, aplicados a la situación de interesado; y (4) las operaciones efectuadas por el tratamiento".

Esta solución nos devuelve a la explicabilidad analizada en la doctrina del Derecho Administrativo por, entre otros, Ponce Solé (2019) para quien los algoritmos y códigos son información pública, o Valero Torrijos (2018) que defiende el derecho de los ciudadanos a obtener información para conocer programas, órgano de control y supervisión, datos empleados, y antecedentes[24]. Gutiérrez David (2021) considera además que es la legislación de transparencia la que debe garantizar la interpretabilidad, explicabilidad y justificación de las decisiones basadas en sistemas ADM. En todo caso, conviene recordar que la explicabilidad de la inteligencia artificial está estrechamente vinculada a la justificación de la decisiones administrativas y dicha exigencia satisface el derecho a la protección judicial efectiva. En definitiva, el control del razonamiento jurídico en vía administrativa y contencioso-administrativa de la actuación de la Administración Pública es el marco normativo en el que operan los sistemas de ADM y por lo tanto, deben responder a los principios que aquí se imponen.

5. LOS LÍMITES DE LA IA EN EL DERECHO

En esta contribución he querido llamar la atención sobre un hecho que suele pasar desapercibido cuando nos plateamos el acomodo los sistemas de ADM en la actividad administrativa: estamos tratando un problema argumentación de las decisiones adoptadas en el marco de una teoría discursiva del Derecho[25]. La moderna teoría de la argumentación fundada por Robert Alexy con *Theorie der juristischen Argumentation*, y por Neil MacCormick con *Legal Reasoning and Legal Theory* es una garantía de la racionalidad del razonamiento jurídico donde se entrelazan la razón teórica que diferencia lo

[23] Decreto No 2017-330, de 14 de marzo de 2017, relativo a los derechos de las personas que sean objeto de decisiones individuales adoptadas sobre el fundamento de un tratamiento algorítmico.

[24] En la doctrina tributaria véase Pérez Bernabeu (2021).

[25] En la Teoría del Derecho, las cuestiones sobre el impacto de la IA en el razonamiento jurídico evocan el debate sobre el papel de la lógica en el Derecho. Para una aproximación brillante a este vasto ámbito teórico véase García Figueroa (2019) y Rodríguez Puerto (2021).

verdadero y lo falso, y la razón práctica que trata con lo que debo hacer como individuo (moralmente) o lo que debemos hacer como comunidad (políticamente) (García Figueroa, 2014, 87), pero también con lo que debemos hacer en el plano altamente institucional del Derecho.

Si aceptamos que el Derecho es argumentación, y que la actividad argumentativa es una actividad discursiva, entonces requiere inexorablemente de seres humanos aportando razones que fundamenten una decisión normativa. Por eso, los sistemas de ADM serán siempre instrumentos que, con limitaciones, puede utilizar el operador jurídico para justificar su decisión (el acto administrativo) y por lo tanto sometidos a control de racionalidad como cualquier otro argumento utilizado en el discurso jurídico. Si los sistemas de ADM se sustraen de la actividad argumentativa, si pensamos que la incorporación de tecnologías de cajas negras logrará emular el razonamiento jurídico, entonces habremos abrazado alguna variante del realismo jurídico, enalteciendo el contexto de descubrimiento[26] y repudiando el paradigma de la racionalidad argumentativa.

La IA y en concreto los sistemas de ADM no son el bálsamo de Fierabrás cervantino, no sirven como solución para todo y menos aún podrían trasladarnos a una situación discursiva ideal. Los sistemas de ADM no vienen a subsanar los límites de la argumentación jurídica, esa actividad discursiva conducida por seres humanos y sujeta a limitaciones de tiempo y conocimiento. Más bien, nos permite reflexionar sobre la esencia de la argumentación jurídica, y sobre las limitaciones que la propia naturaleza del Derecho impone a la IA[27]. No deja de resultar intrigante que la IA se abra paso con tal facilidad cuando buena parte de las posiciones políticas actuales fomentan actitudes y disposiciones incompatibles con ella como la creciente atención a las emociones o la construcción de afectos.

[26] En argumentación jurídica se distingue entre contexto de descubrimiento y contexto de justificación de las decisiones jurídicas. El primero hace referencia a las causas de orden psicológico, sociológico y de otro tipo que determina un acto administrativo o una resolución judicial; el segundo, el contexto de justificación, es el conjunto de razones que se aportan para fundar la decisión. En un Estado de Derecho, la actividad argumentativa discurre en el contexto de justificación porque es ahí donde conocemos las razones que podemos rebatir en sede jurídica. A la Teoría de la Argumentación Jurídica no le interesa ni los motivos personales, ideológicos, sociológicos, psicológicos (García Figueroa, 2014, 142) ni mucho menos datos empíricos o probabilidades de las que ni siquiera se pueden considerar causantes de una decisión.

[27] La llamada de atención la dan Goltz y Gilmore (2018) para quienes el proceso de incorporación de herramientas de IA parece estar liderado por expertos informáticos cuya tendencia normal es explorar todas sus posibilidades y aplicarlas al Derecho. Sin embargo, no todo tiene acomodo ni tiene por qué tenerlo. Los juristas debemos liderar la reflexión sobre cuál es el papel de la IA en el Derecho conociendo los límites de la IA en el Derecho.

Algunos de las limitaciones de la IA en el Derecho ya han sido presentadas anteriormente.

5.1. Las premisas descriptivas no justifican premisas normativas

Las tecnologías que utilizan los sistemas de ADM facilitan premisas descriptivas como patrones de conducta y previsiones que se utilizarán (junto a las premisas normativas como leyes, precedentes, y principios) en la justificación de una decisión jurídica. Como se ha indicado antes, la Ley de Hume es esencial aquí: es imposible derivar enunciados normativos de enunciados descriptivos; por tanto, es imposible derivar una decisión jurídica (juicio de deber) de una razón o premisa descriptiva porque los sistemas de ADM son conceptualmente incapaces de justificar / argumentar / fundamentar decisiones jurídicas (actos administrativos o decisiones judiciales) por sí solos.

5.2. Las máquinas no razonan y menos aún lo hacen jurídicamente

Los programas más complejos de análisis de datos como *Ask Watson a Question* pueden responder o presentar argumentos jurídicos que pueden extraerse de la abrumadora información almacenada; sin embargo, no elaborarán razonamientos jurídicos (Ashley 2017, 3). De hecho, la IA en el Derecho no está modelando la conexión entre el Derecho y la moral, es decir, no pretende encapsular la esencia del razonamiento jurídico como un caso especial de razonamiento práctico[28].

5.3. Las máquinas no son creativas mientras que el razonamiento jurídico sí lo es

La aproximación de los datos en las tecnologías de IA aplicadas al Derecho sí han tenido un efecto disruptivo: las máquinas son realmente eficientes para encontrar correlaciones entre datos y patrones subyacentes, así como probabilidades una vez hecho el análisis cuantitativo de decisiones anteriores. Sin embargo, su rendimiento es deficiente cuando se trabaja con cadenas de razones largas y complejas, ya que éstas dependen de conceptos abstractos, valores, nociones abiertas, principios, políticas, etc. De hecho, Rodríguez Puerto (2021, 83) destaca la inexistencia de modelos de inteligencia artificial que hayan formalizado el proceso de seleccionar y aplicar distintos criterios de interpretación jurídica.

[28] La tesis del caso especial fue propuesta por Robert Alexy y defendida ante las críticas del padre de la teoría discursiva, Jürgen Habermas. Consiste en la integración de los argumentos prácticos generales en el contexto jurídico sin que cambien su carácter, de manera que el la argumentación jurídica es un caso especial del discurso práctico general. Una de las mejores síntesis de esta posición se puede encontrar en Alexy (1999).

Un caso permitirá evidenciar la diferencia entra la solución que aporta un sistema de ADM (las máquinas no razonan) y el razonamiento jurídico de un operador. Cuando la Xunta de Galicia publica ayudas para mujeres emprendedoras en situación de desempleo[29], una mujer desempleada y con domicilio en Galicia solicita una ayuda económica de 5.000 € para poner en marcha una pequeña empresa. La decisión administrativa que se adopta podría estar basada en un sistema de ADM de algoritmo predictivo (similar al utilizado en BOSCO) cuyo resultado es denegar la ayuda porque al acceder a los datos de la Agencia Tributaria, se constata que la solicitante tiene una deuda de 5,88€ y no son elegibles para ayudas públicas las personas deudoras a la Agencia Tributaria (independientemente del monto de la deuda y de la causa que la haya generado).

Ahora bien, cuando se impugna ante la jurisdicción contencioso-administrativa la denegación de la ayuda, el Tribunal Superior de Justicia de Galicia (STSJ 3884/2019) que falló a favor de la demandante, crea el razonamiento jurídico que fundamenta la decisión y entrelaza varias razones:

- La relevancia de las políticas públicas del caso, a saber, la aplicación rigurosa de las normas tributarias y la promoción del empleo y la igualdad.
- El origen de la deuda (insignificante): un recargo del 5% sobre una deuda principal que se pagó previamente.
- El hecho de que el recargo no es autoliquidable, es decir, no se puede pagar hasta que se liquide la deuda principal.
- La inexistencia de deuda cuando la Administración Pública notificó la denegación de la ayuda (la solicitante la liquidó antes de la notificación del acto desestimatorio).

El Tribunal Superior de Justicia de Galicia utilizó argumentos novedosos para socavar el excesivo enfoque formalista en la aplicación de las normas legales. Mientras que la IA aprende repitiendo y extrayendo patrones existentes, la argumentación jurídica requiere creatividad, nuevos argumentos que conecten premisas descriptivas y normativas (principios y valores).

5.4. … Y los límites de la formación jurídica

Sin embargo, esta no es una pugna entre humanos y máquinas porque los sistemas de ADM necesitan a los humanos *in-the-loop*. Plataformas como *Lexis* o *Westlaw* requieren el trabajo minucioso de expertos legales que recopilan y procesan información legal, entrenan la máquina (el algoritmo) y supervisan el

[29] Este interesante caso fue reportado por Nogueira López (2020).

funcionamiento del sistema. Sistemas de ADM como BOSCO o SYRI necesitan traducir al lenguaje informático textos jurídicos. Esto requiere nuevos conjuntos de habilidades y la creación de nuevos roles para los profesionales legales, especialmente en un área, el Derecho, donde la falta de experiencia técnica en el sector jurídico es una barrera funcional clave (Flanagan 2019, 1256). En fin, de nada sirve tener acceso al código fuente de un sistema de ADM si el abogado o incluso el tribunal no tienen capacidad y analizarlo técnicamente para evaluarlo así en términos jurídicos.

6. BIBLIOGRAFÍA

Alexy, Robert (1999) "La tesis del caso especial", en: *Isegoría: Revista de filosofía moral y política* 21, 23-35.

Ashley, Kevin (2017) *Artificial Intelligence and Legal Analytics*, Cambridge University Press, Cambridge.

Boix Palop, Andrés (2020) "Algorithms as Regulations: Considering Algorithms, when used by the Public Administration for Decision-making, as Legal Norms in order to Guarantee the proper Adoption of Administrative Decisions", en: *European Review of Digital Administration & Law* - Erdal 1, 75-99.

Celano, Bruno (1994) *Dialettica della giustificazione pratica: saggio sulla Legge di Humne*, Giappichelli, Torino.

Chen, Daniel L., (2019) "Judicial analytics and the great transformation of American Law", en: *Artificial Intelligence and Law* 27, 15-42.

Coglianese, Cary, Lehr, David (2017) "Regulating by Robot: Administrative Decision Making in the Machine-Learning Era", en: *The Georgetown Law Journal* 105, 1147-1223.

Criado, J. Ignacio (2021) "Inteligencia Artificial (y Administración Pública)", en: *Eunomía. Revista en Cultura de la Legalidad*, 20, 348-372.

Cuéllar, Mariano-Florentino (2016) *Cyberdelegation and the Administrative State,* en: Stanford Public Law Working Paper No. 2754385, http://ssrn.com/abstract=2754385

Flanagan, P., Dewey, M. Hook (2019) "Where do we go from here: Transformation and acceleration of legal analytics in practice", en: *Georgia State University Law Review* 35 (4), 1245-1268.

García Figueroa, Alfonso (2014) "Teoría de la argumentación. Funciones, fines y expectativas", en: Gascón Abellán. M. (ed.), *Argumentación Jurídica,* Tirant lo Blanch, Valencia, 75-110.

— (2017) *Praxis. Una introducción a la moral, la política y el Derecho*, Atelier, Barcelona.

— (2019) *Luís Recaséns Siches. El Jusfilósof demediado (1903-1977)*, en: *80 años del exilio de los juristas españoles acogidos en México*, A. L. López Villaverde (editor), Tirant lo Blanch, Valencia.

García-Pelayo, Manuel (1968) *Del mito y de la razón en la historia del pensamiento político,* Revista de Occidente, Madrid.

Goerlich Peset, José María (2021) "Decisiones administrativas automatizadas en materia social: algoritmos en la gestión de la Seguridad Social y en el procedimiento sancionador", en: *Labos* 2, 22-42.

Goltz, Nachshon, Joel Gilmore (2018) "The Work of Law in the Age of Artificial Intelligence, or How is the Academy Dealing with the "Foruth Revolution?", en: *Robotics, Artificial Intelligence & Law* 1, 27-32.

Gutiérrez David, María Estrella (2021) "Administraciones inteligentes y acceso al código fuente y los algoritmos públicos. Conjurando riesgos de cajas negras decisionales", en: *Derecom* 30m 143-228.

Hagendorff, Thilo (2020), "The Ethics of AI Ethics: An Evaluation of Guidelines", en: *Minds and Machines* 30, 99-120.

Hofman, Herwig C. H. (2021) "An Introduction to Automated Decision-Making and Cyber-Delegation in the Scope of EU Public Law", en: *Univesity of Luxemburg Law Workign Paper Series* 8, 1-12.

Moral Soriano, Leonor (2008), "Precedents: Reasoning by Rules and Reasoning by Principles", en: *Northern Ireland Legal Quarterly* 59, pp. 33-42.

Nogueira López, Alba (2020), "Derechos en la ciudad, vulnerabilidad y derecho a la vivienda", comunicación en el Congreso de la AEPDA, 2020 (Primera sesión): http://www.aepda.es/AEPDAEntrada-2518-XV-CONGRESO-DE-LA-AEPDA.aspx

Pérez Bernabeu, Begoña (2021), "El principio de explicabilidad algorítmica en la normativa tributaria española: hacia un derecho a la explicación individual", en: *Revista española de derecho financiero* 192, 143-178.

Ponce Solé, Juli (2019), "Inteligencia artificial, Derecho administrativo y Reserva de Humanidad: Algoritmos y Procedimiento Administrativo Debido Tecnológico", en: *Revista General de Derecho Administrativo*, 50.

Raso, Jennifer (2021), "AI and Administrative Law", en: Martin-Bariteau, F., y Scassa, T., (eds.) *Artificial Intelligence and the Law in Canada*, LexisNexis, Toronto, 1-17.

Rodríguez Puerto, Manuel (2021), "¿Puede la inteligencia artificial interpretar normas jurídicas? Un problema de razón práctica", en: *Cuadernos Electrónicos de Filosofía del Derecho* 44, 74-96.

Roehl, Ulrik (2022), "Understanding Automated Decision-Making in the Public Sector: A Classification of Automated, Administrative Decision-Making", en: Juell-Skielse, G., Lindgren, I., Åkesson, M. (eds.), *Service Automation in the Public Sector. Progress in IS.* Springer, Cham, 35-63.

Roiblat, HerbertL. (2020), *Algorithms are not enough*, MIT Press, Boston.

Searle, John (1979), *A Taxonomy of Illocutionary Acts*, Cambridge University Press, Cambridge.

— (1983), *Intentionality: An Essay in the Philosophy of Mind*, Cambridge University Press, Cambridge.

Surden, Harr (2019), "Artificial Intelligence and Law: An Overview", en: *Georgia State University Law Review* 35, 1306-1337.

Tashea, Jason (2017), "Courts are Using AI to Sentence Criminals. That must stop now", en: https://www.wired.com/2017/04/courts-using-ai-sentence-criminals-must-stop-now/

Ubaldi, Barbara *et al.* (2019), "State of the art in the use of emerging technologies in the public sector", en: *OECD Working Papers on Public Governance* 34, https://ialab.com.ar/wp-content/uploads/2019/09/OECD-2019-State-of-the-Art-on-Emerging-Technologies-Working-Paper.pdf

Valero Torrijos, Julián (2018), "La tramitación del procedimiento administrativo por medios electrónicos", en Almeida, M. y Míguez, L. (dirs.) *La actualización de la administración electrónica*, Andavira, Santiago de Compostela, 175-216.

Zalnieriute, M., Bennett Moses, L., and Williams, G. (2021), "Automating Government Decision-Making: Implications for the Rule of Law", en S. de Souza, M. Spohr (eds.), *Technology, Innovation and Access to Justice: Dialogues on the Future of Law*, Edinburgh University Press, Edinburgh, 91-111.

CAPÍTULO XX

ESPAÑA DIGITAL 2025. ESTRATEGIA NACIONAL DE INTELIGENCIA ARTIFICIAL

ÁLVARO SÁNCHEZ BRAVO
Universidad de Sevilla

1. INTRODUCCIÓN

Las sociedades tecnológicas tienen ante sí el reto de adaptarse a las nuevas exigencias sociales, culturales, políticas y económicas, con las que están fuertemente imbricadas.

Para ello las nuevas tecnologías, y especialmente Internet, pueden ayudar a la consecución de tales objetivos. Pero no quedándose sólo en la introducción y en el uso de los nuevos recursos tecnológicos, sino asimilando que ello deberá ir inescindiblemente unido a determinados cambios jurídicos, institucionales, organizativos, e incluso de redefinición del propio papel de las administraciones públicas en los nuevos entornos tecnológicos.

Ello supone hacer frente, como premisa, a tres grandes desafíos:

1. Asimilar unas realidades sociales en constante transformación. El sector público es el principal motor de los cambios económicos y sociales, con el horizonte último de una mejora de la calidad de vida de los ciudadanos y de la cohesión social.

Los ámbitos de actuación del sector público, aumentan exponencialmente a medida que aumentan las necesidades de los ciudadanos, pero no sólo de manera cuantitativa, sino cualitativamente. Los ciudadanos no sólo quieren recibir prestaciones de la administración, sino que quieren saber cómo se adoptan las decisiones, e, incluso participar en algunas de ellas. Si la administración está llamada a ser la principal impulsora de los cambios, los ciudadanos quieren ser también protagonistas de esos cambios, y es a través de las nuevas tecnologías, como hoy se hace factible esas fórmulas de participación. Como ya, premonitoriamente, señaló el entonces Comisario Europeo para la Sociedad de la Información: "El sector público, al igual que el resto de la economía, se enfrenta al desafío de reaccionar ante nuevos avances tecnológicos, en concreto en lo que respecta a la tecnología de la información y la comunicación. Internet ha hecho posible nuevas formas de participación en el diseño de políticas, tales como los grupos de opinión en línea, formados de manera muy rápida, o la exigencia a las autoridades públicas de que revisen su modo establecido de tomar decisiones" (Likanen, 2003).

2. Dar respuesta a las nuevas expectativas. Los ciudadanos y las empresas demandan cada vez más, no sólo nuevos servicios, sino que estos se presten de manera más rápida y sin reiterar tramites engorrosos y reiterativos. Todo ello teniendo presente que las necesidades de los ciudadanos son diferentes, y que por lo tanto el objetivo estará en conseguir una atención personalizada al margen de la capacidad, cualificación o ubicación de aquéllos.

3. Ofrecer más con el mismo presupuesto. Las administraciones deben hacer frente a esas nuevas necesidades y exigencias en un escenario de control del gasto público y de contención presupuestaria. Así pues, con el mismo presupuesto deben ofrecerse más y mejores servicios. No olvidemos que debemos desenvolvernos en un entorno de estabilidad presupuestaria pero sin renunciar a las prestaciones de los estados de bienestar como el nuestro.

Ante estos retos, la cuestión parece estar en la consecución de una nueva administración pública que encuentre en las nuevas tecnologías y en la Inteligencia Artificial el motor de su adaptación a las nuevas realidades.

Ahora bien, el salto desde la prestación de determinados servicios, utilizando como soporte los nuevos entornos tecnológicos, hasta la consecución de una disponibilidad generalizada y un uso "universal" de los servicios ofertados exigen que nos detengamos en la consideración de los principales obstáculos que en la hora presente dificultan esa transición. Sin ánimo de ser exhaustivos, dos son los aspectos más relevantes a considerar.

Por un lado, sólo si todos los ciudadanos, en condiciones de igualdad, tienen posibilidad de acceder a los servicios públicos en línea, podrá cimentarse correctamente la administración electrónica.

Uno de los mayores exponentes del desarrollo humano es la posibilidad creciente de los individuos de participar plenamente en la vida social de la colectividad, evitando toda forma de exclusión, con todas las consideraciones políticas y jurídicas que ello implica (Pérez Luño, 2003, especialmente cap. 4).

A ambos objetivos puede contribuir la sociedad de la información e Internet si partimos de la premisa de que es esencial que las personas, o los grupos en que se integran, no sean forzadas a ajustarse a las nuevas tecnologías, si no que sean las nuevas tecnologías las que se adapten a las necesidades de los hombres (Sánchez Bravo, 2010).

Cuestión capital en este ámbito resulta asimismo la consideración de los problemas que las nuevas tecnologías pueden plantear para determinados sectores sociales, tales como los pobres, enfermos, minusválidos, excluidos, marginados, e incluso para los países en vía desarrollo. El riesgo de "fractura digital" vinculado a la desigualdad en el acceso a la tecnología no es

demagogia; es una cuestión real. Esta situación puede ser especialmente difícil para el número considerable de analfabetos que aún existen, para los inmigrantes que desconocen la lengua del país de acogida, y, en general, para aquellas personas que tengan cualquier problema de aprendizaje (Criado Grande, 2001).

Es por ello que los grupos que tengan un riesgo de exclusión sean especialmente integrados en la sociedad de la información, prestando una especial atención a sus concretas necesidades, pues como en la práctica se ha evidenciado, las TIC, y especialmente Internet, pueden contribuir sustancialmente a la mejora de la calidad de vida y la autonomía de numerosas personas que tienen problemas para acceder a determinados servicios o subvenir a sus necesidades empleando los métodos tradicionales.

Desde el punto de vista técnico la inclusión de todos en la sociedad tecnológica requerirá la apuesta por un acceso rápido y generalizado a las nuevas tecnologías. El aumento de usuarios de Internet, de las plataformas digitales, de utilidades tecnológicas, es una realidad constatable, que va unida a un aumento de los contenidos y de la interactividad de los mismos.

Pero para que ello se consolide, y sea verdaderamente operativo, debe producirse la migración a las redes de banda ancha que puedan soportar con fiabilidad y eficacia el evidente aumento del tráfico en la Red, con la generalización de la tecnología 5G. Con todo, los poderes públicos siguen teniendo un papel que desempeñar en los casos en los que los mercados no proporcionan las inversiones necesarias. De este modo, las estrategias nacionales de banda ancha tienen que tener como objetivo incrementar la cobertura en las zonas insuficientemente servidas y estimular la demanda [1].

Por otro lado, un elemento capital en el desarrollo de la sociedad digital es el que el ciudadano confíe en lo que la moderna tecnología le aporta, en que se sienta cómodo y seguro, porque el sistema es fiable y no permite la intrusión de terceros.

Cuestiones como la protección de datos, la certificación digital y la seguridad de los sistemas informáticos son asuntos en los que las administraciones públicas no pueden permitirse ningún fallo. Los ciudadanos tienen que saber que, cuando, para que y durante qué periodo sin datos van a ser usados, almacenados. En definitiva, su derecho a la libertad informática

[1] Infraestructuras digitales seguras y sostenibles. Conectividad: Gigabit para todos, 5G en todas partes. BRÚJULA DIGITAL 2030: SU DÉCADA DIGITAL. COMUNICACIÓN DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO, AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO Y AL COMITÉ DE LAS REGIONES´. Brújula Digital 2030: el enfoque de Europa para el Decenio Digital. COM (2021) 118. Bruselas. 09.03.2021.

debe ser garantizado (Sobre la delimitación del derecho a la libertad informática, vid. Pérez Luño, 1987; 1989; 1989/90; 1992; 1996; 2000; 2001; Lucas Murillo 1989/90; 1990; 1993; 1999; Sánchez Bravo, 1998).

Junto a ello el ciudadano debe tener seguridad en que la Administración podrá comprobar con diligencia y rapidez su identidad, proporcionándole herramientas de autentificación electrónica (firma electrónica y certificados digitales), Así como el desarrollo de aplicaciones para evitar intrusiones no deseadas o "robos de identidad" (Sánchez Bravo, 2001).

Los sistemas electrónicos exigirán, por otro lado, una infraestructura de comunicaciones segura, con unos equipos y programas seguros, que deberán responder a tres exigencias ineludibles: confidencialidad (impedir la divulgación no autorizada de los datos), integridad (impedir la modificación no autorizada de los datos) y disponibilidad (impedir la retención no autorizada de información o recursos) (Sánchez Bravo 1998; 2014).

2. PLAN ESPAÑA DIGITAL 2025

La digitalización tiene potencial para ofrecer soluciones a muchos de los retos a los que se enfrentan Europa y los europeos. Las tecnologías digitales están cambiando no solo la forma en que las personas se comunican, sino también, de manera más general, la manera en que viven y trabajan. Como tal, la pandemia de COVID-19 ha hecho más acuciante la necesidad de acelerar la transición digital en Europa.[2]

Las soluciones digitales contribuyen a la creación de empleo, al progreso de la educación y al aumento de la competitividad y la innovación, y pueden mejorar la vida de los ciudadanos. La tecnología digital tiene un papel clave que desempeñar en la transformación de la economía y la sociedad europeas con el fin de lograr una UE climáticamente neutra de aquí a 2050, objetivo acordado por los dirigentes de la UE.

Salvaguardar los valores de la UE y los derechos fundamentales y la seguridad de los ciudadanos es un elemento clave de la transición digital. La UE pretende seguir un enfoque antropocéntrico que respete las diferencias sociales en toda la Unión.

España, como estado miembro de la UE, participa y colabora en la elaboración e implementación de estas estrategias europeas. Y así, en los últimos años se han ido adoptando diferentes programas, en consonancia con

[2] https://www.consilium.europa.eu/es/policies/a-digital-future-for-europe/#

las agendas digitales europeas[3] [4], mediante la colaboración público-privada y con la participación de todos los agentes económicos y sociales del país. En la elaboración de esta agenda digital han participado más de 15 ministerios y organismos públicos y más de 25 agentes económicos, empresariales y sociales.

España Digital 2025[5] contempla la puesta en marcha durante 2020-2022 de un conjunto de reformas estructurales que movilizarían un importante volumen de inversión pública y privada, en el entorno de los 70.000 millones de euros.

La inversión pública en el periodo 2020-2022 se situaría en torno a los 20.000 millones de euros, de los cuales 15.000 millones de euros, aproximadamente, corresponderían a los diferentes programas y nuevos instrumentos comunitarios de financiación del Plan de Recuperación Next Generation EU[6], que establece que la digitalización tiene que ser uno de los ejes principales para movilizar estos recursos.

A ello se sumaría la inversión prevista por el sector privado, de unos 50.000 millones de euros, en un escenario moderado de despliegue de las medidas.

España Digital 2025 centrará sus objetivos en el impulso a la transformación digital del país como una de las palancas fundamentales para relanzar el crecimiento económico, la reducción de la desigualdad, el aumento de la productividad y el aprovechamiento de todas las oportunidades que brindan las nuevas tecnologías, con respeto a los valores constitucionales y europeos, y la protección de los derechos individuales y colectivos.

Esta agenda consta de cerca de 50 medidas que se articulan en torno a diez ejes estratégicos, entre los cuales merecen destacarse, en la materia que nos ocupa:

---

[3] Comisión Europea, SHAPING EUROPE'S DIGITAL FUTURE. https://ec.europa.eu/info/sites/default/files/communication-shaping-europes-digital-future-feb2020_en_4.pdf

[4] COMUNICACIÓN DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO, AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO Y AL COMITÉ DE LAS REGIONES Brújula Digital 2030: el enfoque de Europa para el Decenio Digital. COM (2021) 118. Bruselas, 09.03.2021.

REGLAMENTO (UE) 2021/694 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 29 de abril de 2021 por el que se establece el Programa Europa Digital y por el que se deroga la Decisión (UE) 2015/2240, DOUE L 166, 11.05.2021.

[5] https://avancedigital.mineco.gob.es/programas-avance-digital/Documents/EspanaDigital_2025_TransicionDigital.pdf

[6] https://ec.europa.eu/info/strategy/recovery-plan-europe_es

2.1. Conectividad digital

Garantizar una conectividad digital adecuada para toda la población, promoviendo la desaparición de la brecha digital entre zonas rurales y urbanas, con el objetivo de que el 100% de la población tenga cobertura de 100 Mbps en 2025.

Deberán fomentarse el uso de redes y servicios digitales, ya que la ir más allá de la disponibilidad de infraestructuras de banda ancha para toda la población. La conectividad entre personas, objetos y empresas sólo existe si las infraestructuras son utilizadas. Es necesario fomentar la utilización de los servicios digitales, comenzando por los usos productivos, buscando apoyo en las fortalezas del sector digital español de servicios de comunicaciones electrónicas, especialmente en términos de identidad digital segura, al objeto de que cualquier persona en cualquier territorio tenga acceso a estos servicios. Para ello, se explorará la posibilidad de desarrollar *bonos de conectividad social* para los colectivos más vulnerables, facilitando así la integración.

Una exigencia vendrá determinada por la transposición de la Directiva 2018/1972, del Parlamento Europeo y del Consejo, de 11 de diciembre de 2018 por la que se establece el Código Europeo de las Comunicaciones Electrónicas[7]. A este respecto, se ha elaborado el Proyecto de Ley de Telecomunicaciones[8],

[7] Directiva (UE) 2018/1972 del Parlamento Europeo y del Consejo, de 11 de diciembre de 2018, por la que se establece el Código Europeo de las Comunicaciones Electrónicas (versión refundida). DOUE L 321, 17.12.2018.

[8] Las modificaciones incorporadas proporcionan mayor seguridad jurídica y flexibilidad a los operadores, mejoran la protección de los derechos de los usuarios y refuerzan las competencias de la Comisión del Mercado de las Telecomunicaciones (CMT). Operadores. El Proyecto de Ley crea un marco de mayor seguridad jurídica e incentivador de las inversiones. En concreto, se crea un marco más adecuado para la realización de inversiones para el despliegue de redes de nueva generación, que permita ofrecer servicios innovadores y tecnológicamente más adecuados a las necesidades de los ciudadanos. Estas redes, tanto fijas como móviles, permitirán ofrecer a los ciudadanos velocidades de acceso a Internet superiores a los 100 Mbits por segundo. Asimismo, establece que la Comisión del Mercado de Telecomunicaciones, a la hora de imponer obligaciones y condiciones de acceso a las redes debe tener en cuenta el riesgo inversor de los operadores. Igualmente, se promueve un uso más eficaz y eficiente del espectro radioeléctrico mediante la generalización de los principios de neutralidad tecnológica (utilización de cualquier tecnología) y de servicios (prestación de cualquier servicio). Respecto a la designación de operador encargado de la prestación del servicio universal, se establece el mecanismo de licitación, mientras que hasta ahora sólo se acudía a este mecanismo si había varios interesados que así lo habían manifestado tras un proceso de consulta. Además, los operadores que pongan su red a disposición de otras entidades para la realización de emisiones radioeléctricas deberán comprobar, previamente al inicio de dichas emisiones, que estas entidades dispongan del correspondiente título habilitante del dominio público radioeléctrico, lo que constituye una importante medida para evitar las emisiones ilegales de radio y televisión. Usuarios. Las modificaciones introducidas refuerzan los derechos de los usuarios y su protección. Así, se establece que los usuarios finales tendrán derecho a recibir mayor información sobre las

(...)

que está en debate, en el momento de escribir este estudio, en sede parlamentaria.

Con el objetivo de convertir a España en un polo de atracción de infraestructuras digitales transfronterizas, tanto puntos de amarre de cables submarino como infraestructuras de almacenamiento y procesamiento de datos, el Gobierno adoptará un Plan de Infraestructuras Transfronterizas.

2.2. Seguir liderando el despliegue de la tecnología 5G en Europa e incentivar su contribución al aumento de la productividad económica, al progreso social y a la vertebración territorial

Se fija como objetivo que en 2025 el 100% del espectro radioeléctrico esté preparado para el 5G. La implementación de servicios e infraestructuras 5G no se confina en el desarrollo de una nueva generación de telefonía móvil, sino que plantean nuevos desafíos, aún no claramente determinados, sobre la transformación industrial y social por sus características de capacidad, baja latencia y densidad de conexiones entre objetos. Estas características técnicas favorecerán usos y nuevos modelos de producción, cambio de las relaciones en las cadenas globales de la economía y el desarrollo de aplicaciones de mayor riqueza en contenidos e interactividad entre personas y objetos, aún en exploración en todo el mundo.

Reforzando la posición de liderazgo de España en el desarrollo y despliegue de redes 5G, es necesario desarrollar un entorno de confianza para el despliegue de estos nuevos servicios, lo que contribuirá a un despliegue temprano de las redes 5G por parte de los operadores económicos.

---

características y condiciones de provisión de los servicios y sobre la calidad con que se prestan (precios, limitaciones de las ofertas, etcétera). También se protegen de modo más eficaz los datos de carácter personal. Por ejemplo, se aplican las normas de protección de datos a aquellos que se obtengan de las etiquetas de los productos comerciales mediante dispositivos de identificación que hacen uso del espectro radioeléctrico (RFID). Además, establece que se debe dar más información al usuario sobre los archivos o programas informáticos ("cookies") que se almacenan en los ordenadores y demás dispositivos empleados para acceder a Internet con el propósito de facilitar la navegación por la red. La nueva normativa precisa que el cambio de operador manteniendo el número (portabilidad) deberá realizarse en el plazo de un día laborable. Asimismo, mejora el acceso a los servicios para personas con discapacidad o con necesidades sociales especiales, estableciendo que deberá ser en condiciones equivalentes al del resto de los usuarios. Con este Proyecto de Ley, esta garantía se extiende a todos los operadores, mientras que antes el acceso sólo se garantizaba para el operador designado para el servicio universal. Por otra parte, se define como infracción el incumplimiento por los operadores de las resoluciones que ponen fin al procedimiento de reclamación de los usuarios. En la Ley de Economía Sostenible se incluye, como parte integrante del servicio universal de telecomunicaciones, que la conexión debe permitir comunicaciones de datos de banda ancha a una velocidad de un Mbit por segundo.

Vid. https://www.lamoncloa.gob.es/consejodeministros/paginas/enlaces/130511-enlaceteleco.aspx

Entre las medidas desarrolladas en este eje estratégico, merece destacarse, y siguiendo lo establecido por la Unión europea[9], el desarrollo de un marco de medidas comunes para mitigar los riesgos de seguridad en las redes 5G, donde se procura mantener un balance entre las medidas de ciberseguridad[10] y el mantenimiento de una competencia efectiva. Sobre esta cuestión volveremos más adelante, al considerar el eje estratégico 4.

2.3. Reforzar las competencias digitales de los trabajadores y del conjunto de la ciudadanía

Como indicamos anteriormente, el desarrollo de programas digitales debe estar abierto para todos, ser inclusivos y no discriminatorios.

Los ciudadanos necesitan competencias digitales, siquiera básicas, para poder desenvolverse con confianza en las redes, comunicarse, informarse o realizar operaciones económicas o entrar en contacto con las administraciones públicas. No obstante los datos evidencian como en España todavía un 43% de la población carece de estas competencias básicas, lo que puede determinar que una amplia capa de población sufra exclusión digital.

Además, se requieren competencias digitales avanzadas para poder desarrollar actividades más avanzadas. Especialmente habrá que considerar la situación de la población activa, donde ya en numerosos sectores, son además necesarias competencias digitales específicas ligadas al trabajo desempeñado, como el manejo de herramientas digitales complejas. *A sensu contrario*, los empleados con competencias digitales limitadas o nulas tienen más riesgo de perder su empleo, acentuándose aún más esta brecha.

Por otro lado, existen sectores estratégicos que trabajan directamente en el mantenimiento y operación de sistemas digitales o en el diseño e implementación de las propias herramientas digitales, lo que crea un nicho de empleo de alto valor añadido, con alta cualificación y salarios.

Por ello, se impone el sistema educativo y la formación a lo largo de toda la vida, asumiendo lo establecido en el Plan Europeo de Educación Digital[11].

---

[9] Recomendación (UE) 2019/534 de la Comisión, de 26 de marzo de 2019, Ciberseguridad de las redes 5G. DOUE L 188. 20.03.2019.

[10] COMUNICACIÓN DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO, AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO Y AL COMITÉ DE LAS REGIONES Despliegue seguro de la 5G en la UE - Aplicación de la caja de herramientas de la UE. COM (2020) 50. Bruselas, 29.01.2020.

[11] (1) el alumnado que actualmente cursa sus estudios primarios o secundarios o estudios de formación profesional debe tener garantías de que adquirirán en el sistema educativo las competencias digitales demandadas por la sociedad para desarrollar una vida plena, personal y laboralmente; (2) la Formación Profesional y la Universidad, junto con las empresas, deberán

(...)

Se contempla, en primer lugar, la implementación del Programa "Educa en Digital"[12], que consiste en un conjunto de acciones para apoyar la Transformación Digital del sistema educativo mediante la dotación de dispositivos, recursos educativos digitales, adecuación de las competencias digitales de los docentes, y acciones que conlleven la aplicación de la Inteligencia Artificial a la educación personalizada.

En segundo lugar, se ha elaborado un Plan Nacional de Competencias Digitales[13], que se vertebra en cuatro ejes de actuación que actúan sobre un conjunto de retos a bordar: (a) la formación digital transversal para la ciudadanía (ciudadanía digital), con énfasis en la capacitación digital de mujeres y niñas, para que todas las personas puedan, entre otras acciones, comunicarse, comprar, realizar transacciones o relacionarse con las Administraciones utilizando las tecnologías digitales con autonomía y suficiencia; (b) el desarrollo de competencias digitales para la educación, desde la digitalización de la escuela hasta la universidad, pasando por la Formación Profesional; (c) la formación en competencias digitales a lo largo de la vida laboral (*upskilling* y *reskilling*, tanto de las personas desempleadas como empleadas), con foco en el desarrollo de competencias digitales para las pymes; y (d) el fomento de los especialistas TIC[14].

Por último, el Plan Uni-Digital se estructura en cuatro líneas estratégicas que se basan en desarrollar proyectos de infraestructuras y servicios TIC, diseñar proyectos de desarrollo de software, generar medidas de apoyo, ayudas e incentivos a la digitalización y la docencia, y un cuarto bloque de medidas estratégicas y de coordinación. De esta manera, algunos de los principales

realizar las adaptaciones necesarias para garantizar que los trabajadores actuales y futuros dispongan de las competencias requeridas; y (3) los agentes y organizaciones sociales, y las Administraciones Públicas deben actuar para incorporar las competencias digitales en la formación a lo largo de la vida. COMUNICACIÓN DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO, AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO Y AL COMITÉ DE LAS REGIONES sobre el Plan de Acción de Educación Digital. COM (2018) 22. Bruselas, 17.01.2018.

[12] Resolución de 7 de julio de 2020, de la Subsecretaría, por la que se publica el Convenio entre el Ministerio de Educación y Formación Profesional, el Ministerio de Asuntos Económicos y Transformación Digital y la Entidad Pública Empresarial Red.es, M.P., para la ejecución del programa "Educa en Digital". BOE 189, 10.07.2020.

[13] Gobierno de España. Plan Nacional de Competencias Digitales. https://portal.mineco.gob.es/RecursosArticulo/mineco/ministerio/ficheros/210127_plan_nacional_de_competencias_digitales.pdf https://portal.mineco.gob.es/RecursosArticulo/mineco/ministerio/ficheros/210127_plan_nacional_de_competencias_digitales.pdf

[14] https://planderecuperacion.gob.es/politicas-y-componentes/componente-19-plan-nacional-de-competencias-digitales-digital-skills

proyectos del plan se centran el desarrollo de infraestructuras de almacenamiento, seguridad y grabación de cursos, entornos de aprendizaje digital y repositorios de código abierto, espacios de interacción y aprendizaje interuniversitario, fortalecimiento del software libre en las universidades, proyectos de formación y medidas de reducción de la brecha digital[15].

2.4. Reforzar la capacidad en ciberseguridad

Los nuevos entornos digitales, amén de numerosas oportunidades, presentan riesgos y amenazas que no son convenientes obviar. Por un lado, los inherentes a los propios incidentes cibernéticos; y por otro, la pérdida de confianza que puede el uso de las nuevas tecnologías, que pueden provocar recelo en los ciudadanos y en los agentes económicos.

A tal efecto, la Estrategia Nacional de Ciberseguridad[16] ha consolidado el hecho de que la ciberseguridad debe ocupar un espacio propio y diferencial, teniendo presente tanto el impacto de la digitalización como motor del cambio con implicaciones para la ciberseguridad más allá del campo meramente de la protección del patrimonio tecnológico para adentrarse en las esferas política, económica y social, como el carácter del ciberespacio como un vector de comunicación estratégica, que puede ser utilizado para influir en la opinión pública y en la forma de pensar de las personas a través de la manipulación de la información, las campañas de desinformación o las acciones de carácter híbrido[17].

[15] https://www.aulamagna.com.es/ministerio-plan-digitalizacion-universidades-espanolas-unidigital/

[16] Orden PCI/487/2019, de 26 de abril, por la que se publica la Estrategia Nacional de Ciberseguridad 2019, aprobada por el Consejo de Seguridad Nacional. BOE 103. 30.04.2019.

[17] La Estrategia se estructura en cinco capítulos.

El capítulo 1 "El ciberespacio como espacio común global" presenta las oportunidades y desafíos del ciberespacio y la infraestructura digital, expone el carácter inherentemente internacional de la aproximación a su seguridad y describe los principales rasgos de la nueva concepción de la ciberseguridad en España.

En el capítulo 2 "Las amenazas y desafíos en el ciberespacio" se examinan las principales amenazas y desafíos del ciberespacio a los que se enfrenta España.

En el capítulo 3 "Propósito, principios y objetivos para la ciberseguridad" se establece el propósito y los principios por los que se rige la Estrategia (unidad de acción, anticipación, eficiencia y resiliencia), así como los objetivos, uno general y cinco específicos que resultan transversales a todos los ámbitos.

- Objetivo general: España garantizará el uso seguro y fiable del ciberespacio, protegiendo los derechos y las libertades de los ciudadanos y promoviendo el progreso socio económico.
- Objetivo I: Seguridad y resiliencia de las redes y los sistemas de información y comunicaciones del sector público y de los servicios esenciales.
- Objetivo II: Uso seguro y fiable del ciberespacio frente a su uso ilícito o malicioso.

(...)

A este respecto, el 3 de mayo de 2022, se ha aprobado un Real Decreto[18], que sustituye al Real Decreto 3/2010, de 8 de enero, por el que se regula el Esquema Nacional de Seguridad en el ámbito de la Administración Electrónica (ENS). El anterior ESN, de 2010, se ha visto superado por la rápida evolución de los contextos normativos, sociales y tecnológicos.

El nuevo ESN establece la política de seguridad para la protección adecuada de la información tratada y los servicios prestados a través de un planteamiento común de principios básicos, requisitos mínimos, medidas de

- Objetivo III: Protección del ecosistema empresarial y social y de los ciudadanos.
- Objetivo IV: Cultura y compromiso con la ciberseguridad y potenciación de las capacidades humanas y tecnológicas.
- Objetivo V: Seguridad del ciberespacio en el ámbito internacional.

En el capítulo 4 "Líneas de acción y medidas" se establecen las líneas de acción dirigidas a la consecución de los objetivos establecidos.

- Línea de Acción 1. Reforzar las capacidades ante las amenazas provenientes del ciberespacio.
- Línea de Acción 2. Garantizar la seguridad y resiliencia de los activos estratégicos para España. Incluye entre sus medidas las siguientes: "3. Asegurar la plena implantación del Esquema Nacional de Seguridad, del Sistema de Protección de las Infraestructuras Críticas, y el cumplimiento y armonización de la normativa sobre protección de infraestructuras críticas y servicios esenciales, con un enfoque prioritario basado en el riesgo." y "5. Desarrollar el Centro de Operaciones de Ciberseguridad de la Administración General del Estado que mejore las capacidades de prevención, detección y respuesta, e impulsar el desarrollo de centros de operaciones de ciberseguridad en el ámbito autonómico y local."
- Línea de Acción 3. Reforzar las capacidades de investigación y persecución de la cibercriminalidad, para garantizar la seguridad ciudadana y la protección de los derechos y libertades en el ciberespacio.
- Línea de Acción 4. Impulsar la ciberseguridad de ciudadanos y empresas
- Línea de Acción 5. Potenciar la industria española de ciberseguridad, y la generación y retención de talento, para el fortalecimiento de la autonomía digital.
- Línea de Acción 6. Contribuir a la seguridad del ciberespacio en el ámbito internacional, promoviendo un ciberespacio abierto, plural, seguro y confiable, en apoyo de los intereses nacionales.
- Línea de Acción 7. Desarrollar una cultura de ciberseguridad.

En el capítulo 5 "La ciberseguridad en el Sistema de Seguridad Nacional" se integra la ciberseguridad en el actual Sistema de Seguridad Nacional con los siguientes componentes:

1. El Consejo de Seguridad Nacional.
2. El Comité de Situación, único para el conjunto del Sistema de Seguridad Nacional ante situaciones de crisis.
3. El Consejo Nacional de Ciberseguridad.
4. La Comisión Permanente de Ciberseguridad.
5. El Foro Nacional de Ciberseguridad.
6. Las Autoridades públicas competentes y los CSIRT de referencia nacionales.

[18] Real Decreto 311/2022, de 3 de mayo, por el que se regula el Esquema Nacional de Seguridad. BOE 106. 04.05.2022.

protección y mecanismos de conformidad y monitorización, para la administración pública, así como los proveedores tecnológicos del sector privado que colaboran con la administración[19].

Para acomodar una respuesta a las amenazas provenientes del ciberespacio, la actualización del ENS persigue tres grandes objetivos.

Primero, alinear el ENS con el marco normativo y el contexto estratégico existentes para garantizar la seguridad en la Administración Digital. Para lograrlo, se clarifica el ámbito de aplicación del ENS y se actualizan las referencias al marco legal vigente, de manera que se simplifiquen y armonicen los mandatos del ENS.

Segundo, introducir la capacidad de ajustar los requisitos del ENS para garantizar su adaptación a la realidad de ciertos colectivos o tipos de sistemas, atendiendo a la semejanza de los riesgos a los que están expuestos sus sistemas de información.

Tercero, reforzar la protección frente a las tendencias en ciberseguridad mediante la revisión de los principios básicos, los requisitos mínimos y las medidas de seguridad que deben adoptarse por las entidades sujetas al ENS.

Los sistemas afectados deberán adecuarse a lo dispuesto en el real decreto en un plazo de veinticuatro meses contados a partir de su entrada en vigor.

El esfuerzo realizado para la actualización del ENS responde a la ejecución de la Reforma 9.3 "Una Administración Cibersegura" del Plan de Digitalización de las Administraciones Públicas 2021-2025, así como a las reformas previstas en la agenda España Digital 2025, con la finalidad de convertirse en una medida urgente de refuerzo del marco normativo en materia de ciberseguridad[20].

---

[19] El Gobierno actualiza el Esquema Nacional de Seguridad en el ámbito de la Administración Pública. Nota de Prensa. MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL. 03.05.2022.

[20] https://administracionelectronica.gob.es/pae_Home/pae_Actualidad/pae_Noticias/Anio2022/Mayo/Noticia-2022-05-04-Publicado-RD-regula-Esquema-Nacional-Seguridad.html

Entre las novedades cabe destacar:

* Se ha revisado y actualizado la redacción del ámbito de aplicación (art 2 y DA 3ª) con una doble finalidad:

En primer lugar, para clarificarlo y que ambos sectores, público y privado (proveedores o suministradores tecnológicos de las entidades del sector público), sean conscientes de lo que les es exigible, en beneficio último de la ciberseguridad pública y de los derechos de los ciudadanos.

(...)

2.5. Impulsar la digitalización de las Administraciones Públicas

La mera existencia de servicios electrónicos no produce eficiencia ni reducción de cargas administrativas, sino que requiere de una modernización de procesos y adaptación de los canales para lograr un uso masivo eficaz, y seguro por ciudadanía y empresas. Por ello, hay margen para mejorar y atender sus demandas, y cumplir el compromiso de excelencia por parte de las Administraciones Públicas[21].

Para ello se hace necesario simplificar las relaciones de los ciudadanos con la administración pública. Esta simplificación permitirá que se puedan personalizar los servicios digitales, que además deben ser fáciles de usar y adaptados, en la medida de lo posible, a las necesidades de cada persona. Garantizando el respeto a la protección de datos personales, se debe minimizar la solicitud de los datos que ya obran en poder de las Administraciones, fomentando la hiperconectividad entre servicios, y se debe permitir personalizar los mecanismos de notificación por los que la ciudadanía opte.

Desde el punto de vista organizativo, se procederá a la integración de todos los niveles administrativos en la transformación digital del sector público[22], con lo que se podrá facilitar la vertebración y cohesión territorial, reducir la brecha digital en la oferta de servicios por parte de la Administración

---

En segundo lugar, para extender su aplicación a los sistemas que manejan o tratan información clasificada, sin perjuicio de que pudiera resultar necesario complementar las medidas de seguridad previstas en el ENS con otras específicas para tales sistemas.

* Se ha realizado la clarificación, precisión, homogeneización, simplificación, o actualización de distintos aspectos del texto, así como la eliminación de aspectos no necesarios o excesivos (un capítulo de 'Comunicaciones electrónicas', con tres artículos, ya superado por las leyes 39/2015, de 1 de octubre y 40/2015, de 1 de octubre, y su desarrollo reglamentario).

* Se han revisado los principios básicos, los requisitos mínimos y las medidas de seguridad:

1. El principio antes denominado 'prevención, reacción y recuperación' pasa a denominarse 'prevención, detección y respuesta', entendiendo que la "recuperación" se encuentra subsumida en el concepto más amplio de "respuesta", que lo incluye.

2. Se introduce el principio de "vigilancia continua" para permitir la detección de actividades o comportamientos anómalos y su oportuna respuesta e impulsar la evaluación permanente del estado de la seguridad de los activos, para detectar vulnerabilidades e identificar deficiencia de configuración.

3. Se clarifica la redacción del principio "responsabilidades diferenciadas" para precisar los aspectos relativos al responsable de la seguridad y al responsable del sistema.

[21] https://portal.mineco.gob.es/RecursosArticulo/mineco/prensa/ficheros/noticias/2018/Agenda_Digital_2025.pdf p.19.

[22] Gobierno de España. Plan de Digitalización de las Administraciones Públicas 2021 -2025 Estrategia en materia de Administración Digital y Servicios Públicos Digitales. https://portal.mineco.gob.es/RecursosArticulo/mineco/ministerio/ficheros/210127_plan_digitalizacion_administraciones_publicas.pdf

General del Estado, las Comunidades Autónomas, y las Entidades Locales, y facilitar la interoperabilidad de los servicios públicos, y en definitiva, facilitar el acceso de la ciudadanía a los servicios, también para aquellos residentes en las áreas de menor densidad de población.

Especial relevancia tendrá la digitalización inteligente que permitirá acompasar la digitalización de la administración con los avances tecnológicos. A este respecto, el impulso a la aplicación de servicios de automatización, capacidades de Inteligencia Artificial reutilizables y servicios de gestión inteligentes, facilitará una transformación efectiva de los procesos de articulación y ejecución de las políticas públicas, simplificando y automatizando los procesos que resulten en un mayor bienestar para la ciudadanía y en una mayor eficiencia empresarial.

El Plan de Digitalización de las Administraciones Públicas (PDAP)[23], pretende un salto decisivo en la mejora de la eficacia y eficiencia de la Administración Pública, buscando dar respuesta a los retos de los principales ámbitos tractores de la Administración Pública, como son el empleo, la justicia y la sanidad, y tiene por objeto mejorar la eficiencia de las Administraciones Públicas en su conjunto, garantizando la sostenibilidad de las inversiones mediante el refuerzo y reutilización de medios y servicios compartidos.

Sus tres objetivos fundamentales pueden agruparse en:

a) Servicios digitales, accesibles, eficientes, seguros y fiables: Desarrollar servicios públicos digitales más inclusivos, eficientes, personalizados, proactivos y de calidad para para el conjunto de la ciudadanía.

b) Políticas públicas basadas en datos y modernización de la gestión de datos: Transformar a la Administración Pública española en una Administración más moderna y "guiada por datos", donde la información de los de los ciudadanos, las ciudadanas y de las Administraciones Públicas se utiliza eficientemente para diseñar políticas públicas alineadas con la realidad social, económica y territorial de España, así como para la construcción de una experiencia ciudadana de los servicios públicos verdaderamente innovadora.

c) Democratización del acceso a las tecnologías emergente: Permitir desarrollar servicios, activos e infraestructuras comunes que permitan a todas las Administraciones Públicas sumarse a la

[23] Ibíd.

revolución tecnológica que está suponiendo la irrupción de nuevos habilitadores tecnológicos como pueden ser la Inteligencia Artificial o la tecnología de analítica de datos.

2.6. Garantizar los derechos en el nuevo entorno digital

Como señalamos anteriormente, los nuevos escenarios digitales están modificando profundamente nuestras realidades y nuestros modelos sociales, desde la forma en que consumimos, nos relacionamos con otros e trabajamos. Pero la revolución tecnológica se haya en el albur de su evolución, con un desarrollo tecnológico exponencial que corre más rápido, en numerosas ocasiones, que el tiempo que necesitamos para valorar sus aportes o riesgos. La pregunta es, por tanto, si nuestros marcos jurídicos y éticos son conformes ante estos nuevos apremios, y consecuentemente, la necesidad de ponernos en marcha hacia nuevos marcos regulatorios.

En este sentido, el primer objetivo específico lo constituye la elaboración de una Carta de Derechos Digitales, que delimite los derechos de ciudadanos y operadores en el mundo digital estableciendo un marco de certidumbre sobre la hermenéutica de determinados principios, y reconociendo nuevos derechos acordes a los nuevos escenarios tecnológicos. Además, como señala, el propio Plan, contribuirá a reducir las brechas digitales que se han ampliado, en los últimos años, por motivos socioeconómicos, de género, generacionales o territoriales. En concreto, la implementación del derecho de acceso a Internet de calidad y asequible en todo el territorio nacional, así como a la formación, capacitación y desarrollo de habilidades digitales en todos los sectores de la población, especialmente entre los colectivos más vulnerables, serán claves para luchar contra las brechas digitales y permitir la articulación territorial del país.

Esta certidumbre proporcionará un escenario adecuado para un desarrollo de nuevos productos y servicios basados en tecnologías digitales, eliminando así la creencia, aún mantenida en algunos sectores, que en la sociedad digital todo era posible, y que podría hacer en la red lo que no está permitido en el mundo físico; o viceversa.

En un principio la desregulación constituyó un aliciente para fomentar la innovación, pero con unos sistemas digitales avanzados, y con experiencia suficiente para perfilar las ventajas y riesgo de su uso, se hace necesario, cuando no urgente, un marco regulatorio que de seguridad jurídica prescribiendo lo que está autorizado y lo que no.

El 14 de julio de 2021 se presentó la Carta de Derechos Digitales[24], que como indica el propio texto, "no tiene carácter normativo", sino que propone "un marco de referencia para la acción de todos los poderes públicos, que, siendo compartido por todos, permita navegar el entorno digital aprovechando y desarrollando todas sus potencialidades y oportunidades". Además, "pretende servir de guía para futuros proyectos legislativos y desarrollar políticas públicas más justas, que nos protejan a todos"[25].

A mayor abundamiento, se explicita claramente como "no se trata necesariamente de descubrir derechos digitales pretendiendo que sean algo distinto de los derechos fundamentales ya reconocidos o de que las nuevas tecnologías y el ecosistema digital se erijan por definición en fuente de nuevos derechos. La persona y su dignidad son la fuente permanente y única de los mismos y la clave de bóveda tanto para proyectar el Ordenamiento vigente sobre la realidad tecnológica, como para que los poderes públicos definan normas y políticas públicas ordenadas a su garantía y promoción…la Carta de derechos digitales que se presenta no trata de crear nuevos derechos fundamentales sino de perfilar los más relevantes en el entorno y los espacios digitales o describir derechos instrumentales o auxiliares de los primeros. Se trata de un proceso naturalmente dinámico dado que el entorno digital se encuentra en constante evolución con consecuencias y límites que no es fácil predecir"[26].

Los derechos se articulan en torno a cinco grandes principios:

Respecto a los derechos de libertad, el texto incluye el derecho a la identidad del entorno digital, a la protección de datos[27], al pseudonimato, el derecho a no ser localizado y perfilado, el derecho a la ciberseguridad, o el derecho a la herencia digital.

En cuanto a los derechos de igualdad, la Carta recoge el derecho a la igualdad y a la no discriminación en el entorno digital, el derecho de acceso a Internet y el derecho de accesibilidad universal en el entorno digital.

El texto también promueve la protección de menores en el entorno digital para que tutores o progenitores velen porque los menores de edad hagan un uso equilibrado de entornos digitales, garanticen el adecuado desarrollo de su personalidad y preserven su dignidad; además promueve el fomento del

[24] https://www.lamoncloa.gob.es/presidente/actividades/Documents/2021/140721-Carta_Derechos_Digitales_RedEs.pdf

[25] https://www.lamoncloa.gob.es/presidente/actividades/Paginas/2021/140721-derechos-digitales.aspx

[26] Ibíd.

[27] Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. BOE 294. O6.12.2018. Especialmente, Título X.

acceso a todos los colectivos y la promoción de políticas públicas para eliminar brechas de acceso al entorno digital.

El derecho a la neutralidad de la red, a recibir libremente información veraz, el derecho a la participación ciudadana por medios digitales y el derecho a la educación digital son otras de las novedades del texto en el apartado de derechos de participación y conformación del espacio público.

En el ámbito laboral, la Carta de Derechos Digitales recoge el derecho a la desconexión digital, al descanso y a la conciliación de la vida personal y familiar, la evaluación de impacto en el uso de los algoritmos o el desarrollo de condiciones óptimas para la creación de espacios de pruebas controladas (sandbox).

En relación con los derechos en entornos específicos, se incluyen contenidos muy novedosos y pioneros. Es el caso los derechos ante la inteligencia artificial. El texto recoge que la IA deberá asegurar un enfoque centrado en las personas y su inalienable dignidad y que en el desarrollo de los sistemas de inteligencia artificial se deberá garantizar el derecho a la no discriminación. También se incluyen los derechos digitales en el empleo de las neurotecnologías para, entre otras cuestiones, garantizar el control de cada personal sobre su propia identidad, asegurar la confidencialidad y asegurar que las decisiones y procesos basados en estas tecnologías no sean condicionados por el suministro de datos[28].

Como garantías se establecen, entre otras, el derecho de todas las personas a la tutela administrativa y judicial de sus derechos en los entornos digitales.

Especial importancia debía prestarse a la regulación normativa aplicable al trabajo a distancia, estableciendo una normación *ad hoc* que garantizara la seguridad jurídica y que permita el desarrollo de esta modalidad laboral, como medio de organización empresarial que promueva la productividad y la eficiencia, garantizando la protección de los derechos de los trabajadores ante los nuevos entornos digitales.

Esta previsión se ha visto colmada con la aprobación de la Ley de Trabajo a Distancia[29], que nace con el objetivo de proporcionar una regulación suficiente, transversal e integrada en una norma sustantiva única que dé respuestas a diversas necesidades, equilibrando el uso de estas nuevas formas de prestación de trabajo por cuenta ajena y las ventajas que suponen para

[28] https://administracionelectronica.gob.es/pae_Home/pae_Actualidad/pae_Noticias/Anio2021/Julio/Noticia-2021-07-15-El-Gobierno-de-Espana-adopta-Carta-Derechos-Digitales.html

[29] Ley 10/2021, de 9 de julio, de trabajo a distancia. BOE 164. 10.07.2021.

empresas y personas trabajadoras, de un lado, y un marco de derechos que satisfagan, entre otros, los principios sobre su carácter voluntario y reversible, el principio de igualdad de trato en las condiciones profesionales, en especial la retribución incluida la compensación de gastos, la promoción y la formación profesional, el ejercicio de derechos colectivos, los tiempos máximos de trabajo y los tiempos mínimos de descanso, la igualdad de oportunidades en el territorio, la distribución flexible del tiempo de trabajo, así como los aspectos preventivos relacionados básicamente con la fatiga física y mental, el uso de pantallas de visualización de datos y los riesgos de aislamiento.

2.7. Transitar hacia una economía del dato, garantizando la seguridad y privacidad y aprovechando las oportunidades que ofrece la Inteligencia Artificial

El *big data* está en el centro de todas las grandes transformaciones que la digitalización tiene en la sociedad actual. Los datos pueden ser creados por personas o generados por máquinas, como sensores que recopilan información climática, imágenes de satélite, fotografías y videos digitales, registros de transacciones de compra, señales de GPS y más. Cubre muchos sectores, desde la salud hasta el transporte y la energía.

La generación de valor en las diferentes etapas de la cadena de valor de los datos estará en el centro de la futura economía del conocimiento. El buen uso de los datos también puede brindar oportunidades a sectores más tradicionales como el transporte, la salud o la manufactura[30].

Ahora bien, debe darse prioridad a las personas en el desarrollo tecnológico y promoverlos valores y derechos democráticos en el mundo digital[31]. El uso extendido y la gestión de los datos mediante la acción de los algoritmos y sistemas autónomos tienen múltiples implicaciones en el plano ético y moral que exige procesos y mecanismos de control que protejan nuestros valores, principios y derechos.

Igualmente, se establecen otros, como aumentar la digitalización de las empresas, con especial atención a las micropymes y a las start-ups, acelerar la digitalización del modelo productivo mediante proyectos tractores de transformación digital en sectores económicos estratégicos como el Agroalimentario, Movilidad, Salud, Turismo, Comercio o Energía, entre otros; y

[30] https://digital-strategy.ec.europa.eu/en/policies/big-data

[31] Propuesta de REGLAMENTO DEL PARLAMENTO EUROPEO Y DEL CONSEJO, sobre gobernanza europea de datos (Ley de gobernanza de datos), COM (2020) 767. Bruselas, 25.11.2020.

mejorar el atractivo de España como plataforma audiovisual europea para generar negocio y puestos de trabajo.

Además, España Digital 2025 quiere contribuir a cerrar las diferentes brechas digitales que se han ensanchado en los últimos años, ya sea por motivos socioeconómicos, de género, generacionales, territoriales, o medioambientales, y que se han puesto de manifiesto durante la pandemia. Una misión que se encuentra alineada a los Objetivos de Desarrollo Sostenible (ODS) y la Agenda 2030 de Naciones Unidas.

3. ESTRATEGIA NACIONAL DE INTELIGENCIA ARTIFICIAL

La Inteligencia Artificial (IA) se está desarrollando a alta velocidad en todo el planeta, debido en parte, a la gran proliferación de datos. Sin embargo, su extremada versatilidad es también una fuente potencial de riesgos (discriminación provocada por conjuntos de datos sesgados; decisiones automatizadas difíciles de entender; intrusión en la vida privada de las personas; o utilización con propósitos delictivos) si no se respetan determinadas reglas (Sánchez Bravo, 2020).

El aumento de la capacidad computacional hizo que fuera posible la implementación de algoritmos cada vez más complejos, potentes y flexibles. Al mismo tiempo, la amplia disponibilidad de datos dio lugar a grandes avances en el campo de la inteligencia artificial (IA). Los datos están, por tanto, en el centro de esta transformación. Pero la forma en que se recojan y utilicen los datos debe situar los intereses de las personas en primer lugar, conforme los valores, derechos fundamentales las normas jurídicas propias de Estados democráticos de Derecho.

Uno de los más relevantes objetivos de los sistemas modernos de IA es distinguir y extraer patrones de datos sin procesar para construir su propio conocimiento. Frente a los sistemas expertos, la solución actual no es trabajar con una base de datos de conocimiento, sino aprender conocimiento. Esa capacidad de la IA para aprender se conoce como aprendizaje de máquina y permiten que las computadoras resuelvan problemas que requieren cierta comprensión del mundo real y tomen decisiones situacionales y subjetivas.

El aprendizaje supone que las máquinas puedan encontrar patrones diferentes de los generalmente asimilados por los cerebros humanos[32].[33]

[32] Comisión Económica para América Latina y el Caribe (CEPAL), Datos, algoritmos y políticas: la redefinición del mundo digital (LC/CMSI.6/4), Santiago, 2018, p. 169-176.

[33] La IA puede utilizarse para el desarrollo económico y social, basado en los Objetivos de Desarrollo del Milenio (ODM), identificándose cuatro elementos que enmarcan los efectos de la IA en el desarrollo:

(...)

El crecimiento de la capacidad informática y la disponibilidad de datos, así como los avances de los algoritmos hacen de la inteligencia artificial una de las tecnologías más estratégicas del siglo XXI.

Transferencias de inteligencia.

1) La inteligencia a distancia hace referencia a que las modernas redes de telecomunicaciones permiten aplicar a distancia sistemas de inteligencia artificial altamente entrenados.

La inteligencia a distancia es la capacidad de las tecnologías de inteligencia artificial (IA), en combinación con las telecomunicaciones, para remediar la carencia de recursos en campos que no cuentan con personal suficiente o han sido poco investigados. Esto es especialmente importante si se considera que el aprendizaje multitarea y de transferencia permite reutilizar la inteligencia generada u obtenida en otro lugar. Una de las aplicaciones pioneras es el uso de la IA en los sectores de educación y salud, como en el caso de la educación a distancia automatizada y los diagnósticos a distancia para tratar una serie de enfermedades (cataratas congénitas, tuberculosis y cáncer de mama, entre otras).

En el ámbito de la educación, las soluciones de IA permiten automatizar los sistemas de educación y tutoría, proceso que a su vez permite soluciones de bajo costo a gran escala. Se pueden automatizar actividades especialmente estructuradas, como el aprendizaje de idiomas, la programación de software o las habilidades analíticas cuantitativas. Los sistemas de IA de aprendizaje posibilitan la masificación de una experiencia de educación individualizada para un curso estructurado.

La inteligencia a distancia puede revolucionar la industria de la salud al incrementar la eficiencia y la cobertura.

2) La inteligencia local se refiere al hecho de que los sistemas IA se pueden aplicar de forma autónoma localmente, adaptándose al contexto y requisitos locales.

Casos emblemáticos son los relacionados con el cambio climático, igualdad de género en el lugar de trabajo y en el aula y ciudades inteligentes.

Manipulación de la realidad.

3) La realidad aumentada, virtual y duplicada se refiera al hecho de que la IA permite crear los llamados *gemelos digitales* de aspectos de la realidad que luego puedan usarse para mejorar nuestra comprensión de la realidad o duplicar aspectos ésta.

En muchas aplicaciones prácticas, la inteligencia a distancia y local se combina cada vez más con el uso de la realidad virtual y aumentada. Así los vehículos autónomos, por ejemplo, pueden usar mapas tridimensionales para tomar decisiones en tiempo real. Las realidades virtuales guiadas por la IA también se utilizan para fomentar la educación y la igualdad de género.

Más allá de las realidades aumentadas y virtuales, la IA también se está utilizando para duplicar el diseño de átomos del mundo real y objetos moleculares, como los alimentos. Se trata de duplicar la estructura de un determinado artículo para desarrollar una versión más sostenible de este. La duplicación se podría utilizar para combatir el hambre.

4) La realidad de grano fino hace referencia al hecho de que la huella digital proporciona mapas cada vez más detallados de la realidad y el aprendizaje de máquina permite explotar la información resultante para impulsar el logro de los objetivos de desarrollo.

Una de las formas en que la IA puede proporcionar información más detallada sobre áreas específicas en materia de desarrollo económico y social es refinando nuestra comprensión de la realidad mediante una nueva manera de recopilar datos con mayor granularidad. El aprendizaje automatizado de representación permite transformar detalles recién obtenidos en características útiles.

Ibídem.

A este respecto, y como medida fundamental, entre otras, se ha elaborado la Estrategia Nacional de Inteligencia Artificial (ENIA)[34].

Por IA debe entenderse, conforme a los señalado por la Comisión Europea, "el software que se desarrolla empleando una o varias de las técnicas y estrategias que figuran en el anexo I y que puede, para un conjunto determinado de objetivos definidos por seres humanos, generar información de salida como contenidos, predicciones, recomendaciones o decisiones que influyan en los entornos con los que interactúa"[35].

Como señala ENIA, por una serie de factores, que parten de sistema de programación humana y esquemas de decisión predeterminados, se ha puesto en marcha un proceso sin precedentes, y parece que irreversible, de expansión de la IA que está modificando nuestros sistemas económicos, sociales y jurídicos, y que encuentra su apoyo en determinados elementos configuradores de esta nueva realidad: El enorme crecimiento en la cantidad de datos disponibles; los avances en la potencia y capacidad de los sistemas de computación y almacenamiento; y la investigación y desarrollo con éxito de nuevos algoritmos y métodos de aprendizaje automático[36].

La situación en nuestro país respecto a la IA ha mejorado considerablemente en los últimos años, con relevantes inversiones en infraestructuras y tecnologías de la información y la comunicación, pero todavía existen relevantes retos que es necesario considerar: Así, es necesario avanzar en los siguientes sectores estratégicos:

- Aumentar las competencias digitales de la población, en especial la de las personas en situación o riesgo de exclusión social.

---

[34] Estrategia Nacional de Inteligencia Artificial 2020. https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/ENIA2B.pdf

[35] Propuesta de REGLAMENTO DEL PARLAMENTO EUROPEO Y DEL CONSEJO POR EL QUE SE ESTABLECEN NORMAS ARMONIZADAS EN MATERIA DE INTELIGENCIA ARTIFICIAL (LEY DE INTELIGENCIA ARTIFICIAL) Y SE MODIFICAN DETERMINADOS ACTOS LEGISLATIVOS DE LA UNIÓN. COM (2021) 206. 21.04.2021. Art. 3.1.

Por su parte, el Anexo I establece: "ANEXO I. TÉCNICAS Y ESTRATEGIAS DE INTELIGENCIA ARTIFICIAL mencionados en el artículo 3, punto 1:

Estrategias de aprendizaje automático, incluidos el aprendizaje supervisado, el no supervisado y el realizado por refuerzo, que emplean una amplia variedad de métodos, entre ellos el aprendizaje profundo.

Estrategias basadas en la lógica y el conocimiento, especialmente la representación del conocimiento, la programación (lógica) inductiva, las bases de conocimiento, los motores de inferencia y deducción, los sistemas expertos y de razonamiento (simbólico).

Estrategias estadísticas, estimación bayesiana, métodos de búsqueda y optimización.

[36] Estrategia Nacional de Inteligencia Artificial 2020, p. 2. https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/ENIAResumen2B.pdf p.2

- Acelerar la digitalización del tejido de pequeñas y medianas empresas (PYMEs).
- Promover la creación de repositorios de datos y facilitar el acceso a los mismos.
- Mejorar la eficiencia y productividad de los servicios públicos.
- Estimular la colaboración e incrementar la inversión pública y privada en I+D+I. [37].

A este respecto, el plan estatal resalta la capital importancia que se otorga a los poderes públicos y su liderazgo, pues "en la medida en que contribuye a poner el desarrollo tecnológico al servicio de la sociedad y como un factor de salvaguarda de nuestro estado de bienestar social. Sanidad, Educación, Justicia, dependencia y sistema de prestaciones son pilares y marca distintiva de España como país. España busca que la IA contribuya a consolidar nuestro estado de bienestar, aportando a su vez los datos y activos necesarios para impulsar la innovación y un desarrollo tecnológico por y para la sociedad, en un círculo virtuoso entre la tecnología y nuestro sistema político, social, económico e industrial"[38].

Para ello, ENIA se diseña, no sólo como una estrategia de investigación científica, y/o un campo prioritario de innovación empresarial y desarrollo industrial, sino que pretende convertir a la IA en el gran vector de cambios estructurales en la sociedad española.

Para ello, se requiere, siguiendo lo explicitado en la Estrategia, "una aproximación interdisciplinar centrada en las personas y el medio ambiente, que incorpore las distintas perspectivas de la Ingeniería en Informática, las ingenierías técnicas, las matemáticas, la biología, la neurociencia, la sociología, psicología, la economía, la física, las ciencias terrestres y ambientales, el derecho y las humanidades, con el fin de impulsar el despliegue de la IA en un marco que preserve nuestros valores democráticos, y el respeto al marco de derechos individuales y colectivos"[39].

ENIA se articula en torno a siete Objetivos Estratégicos (OE):

- Excelencia científica e innovación en Inteligencia Artificial. Situar a España como país comprometido a potenciar la excelencia científica y la innovación en Inteligencia Artificial.

[37] Ídem, p. 3.

[38] Ibídem .

[39] Estrategia Nacional de Inteligencia Artificial 2020, cit., p. 10.

- Proyección de la lengua española. Liderar a nivel mundial el desarrollo de herramientas, tecnologías y aplicaciones para la proyección y uso de la lengua española en los ámbitos de aplicación de la IA.
- Creación de empleo cualificado. Promover la creación de empleo cualificado, impulsando la formación y educación, estimulando el talento español y atrayendo el talento global.
- Transformación del tejido productivo. Incorporar la IA como factor de mejora de la productividad de la empresa española, de la eficacia en la Administración Pública, y como motor del crecimiento económico sostenible e inclusivo.
- Entorno de confianza en relación a la Inteligencia Artificial. Generar un entorno de confianza en relación a la IA, tanto en el plano de su desarrollo tecnológico, como en el regulatorio y en el de su impacto social.
- Valores humanistas en la Inteligencia Artificial. Impulsar el debate a nivel global sobre el desarrollo tecnológico de valores humanistas (Human-Centered AI), centrado en velar por el bienestar de la sociedad a la hora de realizar avances o desarrollos tecnológicos, creando y participando en foros y actividades divulgativas para el desarrollo de un marco ético que garantice los derechos individuales y colectivos de la ciudadanía.
- Inteligencia Artificial inclusiva y sostenible. Potenciar la IA inclusiva y sostenible, como vector transversal para afrontar los grandes desafíos de nuestra sociedad, específicamente para reducir la brecha de género, la brecha digital, apoyar la transición ecológica y la vertebración territorial.

En la misma línea incorpora la necesidad de que el diseño de estos sistemas sea robusto, seguro e imparcial, para avanzar hacia una IA fiable, explicable, transparente e inclusiva que asegure el cumplimiento de los derechos fundamentales y de la regulación aplicable, así como el respeto a los principios y valores fundamentales, y tenga en cuenta las aspiraciones colectivas de la ciudadanía[40].

[40] https://www.boe.es/boe/dias/1999/10/20/pdfs/A36825-36830.pdf en el que se dispuso que los derechos fundamentales son el fundamento básico para garantizar la "primacía del ser humano" en un contexto de cambio tecnológico, y es de manera similar, como se ha propuesto la "Guía de ética de AI" https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai producida por el Grupo de expertos de alto nivel sobre IA de la Comisión Europea.

Estos objetivos se articularán, en el período 2020-2025, en seis Ejes de Actuación Estratégicos (EAE), que definirán las líneas de actuación en cada uno los EAE.

- *Impulso de la investigación científica, desarrollo tecnológico e innovación en IA.* Entre las medidas dirigidas a generar este impulso, se anuncia la Red Española de Excelencia en IA, un programa de ayudas a empresas para el desarrollo de soluciones en IA y datos, la creación de un Programa de Misiones de I+D+i en IA o el refuerzo de la red de Centros de Innovación Digital.
- *Fomento de las capacidades digitales, desarrollo del talento nacional y atracción del internacional.* En este punto, se enuncian medidas tales como el desarrollo del Plan Nacional de Competencias Digitales o promover la formación a través de diversos programas.
- *Desarrollo de plataformas de datos e infraestructuras tecnológicas que den soporte a la IA*. Con esta finalidad, se informa de la creación de la *Oficina del Dato y del Chief Data Officer* (un nuevo rol sobre el que cada vez escuchamos hablar más, sobre todo en otros países de la Unión Europea), o la creación de espacios compartidos de datos sectoriales e industriales y repositorios descentralizados y accesibles (que sigue la línea de la Estrategia de Datos de la Unión Europea).
- *Integración de la IA en las cadenas de valor para transformar el tejido empresarial*. Para conseguir este objetivo, se prevé el lanzamiento de programas de ayudas para empresas para la incorporación de IA en los procesos productivos de las cadenas de valor, programas de impulso a la transferencia de innovación en IA (veremos a ver si esto se traduce en un mayor movimiento en las transacciones tecnológicas) o el lanzamiento de un fondo (fondo "*NextTech*") de capital riesgo público-privado para impulsar el emprendimiento digital y crecimiento de empresas en IA.
- *Impulso del uso de la IA en la Administración Pública y en las misiones estratégicas nacionales*, a través del fomento de las competencias en IA y su incorporación a la administración pública.
- *Establecimiento de un marco ético y normativo* que garantice la protección de los derechos individuales y colectivos, con el bienestar social y la sostenibilidad. En este eje, se enmarca, entre otras cuestiones, el desarrollo de un sello nacional de calidad IA o el desarrollo de la Carta de Derechos Digitales o la puesta en marcha de un modelo de gobernanza nacional de la ética en la IA.

Esta somera exposición nos permite trazar las líneas fundamentales de esta relevante iniciativa patria, que como señala la propia ENIA, y ante la situación de crisis mundial, virus y guerra, la IA está jugando un papel primordial en las soluciones implementadas y las que están por llegar.

Es muy importante aprender las lecciones sobre los principales problemas que se han puesto sobre la mesa, para tomar las medidas necesarias para corregir las disfuncionalidades, avanzar en nuevas medidas, y preparar los sistemas tecnológicos para para en el futuro puedan obtener aún mayor partido de la aplicación de la tecnología digital en general y de la IA en particular a gestiones de posibles, aunque no deseadas, situaciones de crisis como la actual.

Concluyamos con las propias reflexiones de la ENIA, al señalar las lecciones aprendidas:

"1. Contar con mecanismos que aseguren la recolección de datos de calidad y estandarizados. 2. Tomar siempre en consideración todos los aspectos relacionados con el uso de estos sistemas (privacidad y seguridad de datos, accesibilidad, usabilidad) 3. Asegurar que la IA se aplica sólo cuando es realmente necesaria y no despertar expectativas exageradas 4. Establecer metodologías para la evaluación formal y medición de resultados tras la aplicación de sistemas IA. 5. Incrementar la cooperación internacional y entre áreas de conocimiento (ciencias, tecnologías, humanidades, etc.) en la búsqueda de soluciones a este tipo de crisis".[41]

4. A MODO DE CONCLUSIÓN

Las nuevas tecnologías digitales son ya una realidad. Instaladas en la práctica cotidiana y la consciencia tecnológica de nuestras sociedades su crecimiento es ciertamente exponencial. Pero ante todo creo necesario una última reflexión sobre la percepción de los fenómenos tecnológicos por los ciudadanos y el papel que en el mismo corresponde a las instancias públicas.

En esta nueva sociedad digital, el ciudadano debe seguir siendo el centro de toda política, el alfa y el omega, el principio y el fin. Lo contrario es arrastrar los procesos tecnológicos a una despersonalización que los haga asfixiantes e inoperantes frente a los ciudadanos. La sociedad digital que se pretende implementar debe tener en el flujo de informaciones el instrumento indispensable para que los distintos pueblos y sociedades, se conozcan, se integren y comportan ideales e inspiraciones comunes. El intercambio de información, de informaciones de calidad (big data cualificado), es la base sobre la que fomentar el respeto mutuo, la tolerancia y la consecución de metas comunes, sin renuncia a la idiosincrasia propia. La información puede

[41] Estrategia Nacional de Inteligencia Artificial 2020, cit., p. 84.

consolidarse como un poderoso instrumento para acabar con las finiseculares y tópicas concepciones de unos contra otros, fruto en muchos casos de la desinformación y la ignorancia. El mundo se compone de numerosos pueblos, cada uno con su lengua, su cultura y sus modos de expresión. Las TIC y la digitalización, pueden usarse, bien para aumentar las diferencias entre unos y otros, o bien para integrarnos a todos desde el mutuo conocimiento y respeto.

Al ser conscientes de las múltiples implicaciones de los procesos tecnológicos, se debe articular un importante esfuerzo por implementar un desarrollo tecnológico que no aniquile su componente subjetiva.

Determinante es también, al margen de las iniciativas públicas o privadas de promoción, la consideración que los ciudadanos tienen de los procesos tecnológicos y de los beneficios y/o perjuicios que acarrearán para sus formas de vida, de pensamiento y de entender el mundo que les rodea. Internet es un caso paradigmático del proceso que anunciamos. Panacea salvadora para unos; "Caja de Pandora" para otros.

De esta forma, observamos como bajo la imperiosa necesidad de "digitalizarse" el número de adeptos incondicionales al proceso de digitalización se multiplica cada vez más. Para ellos la informatización es la solución, que puede resolver de una vez para todas y para siempre todos los males que aquejan a la sociedad actual. Enfrente, los detractores del sistema, los que auguran la aniquilación del hombre, la toma del poder por las maquinas, la supremacía de la inteligencia artificial sobre la inteligencia humana (Pérez Luño, 1986-87; Madrid Conesa, 1984).

Como gráfica y acertadamente ha señalado Pérez Luño, no se trata de subirse al carro de los apocalípticos o de los integrados, sino de someter la utilización de la informática a unas garantías jurídicas (Pérez Luño, 1986/87).

Por su parte, la situación excepcional derivada de la pandemia de la COVID-19 ha acelerado el proceso de digitalización, poniendo de relieve las fortalezas y también las carencias tanto desde el punto de vista económico como social y territorial. Por ello, deben abordarse de manera urgente los apremios aún no resueltos que permitan articula una sociedad digital inclusiva, que asegure la accesibilidad del conjunto de la sociedad a las oportunidades de la sociedad digital.

Todo ello explica la elaboración y determinación que el Plan 2025 España supone, para, como señala el propio Plan, "articular una Agenda actualizada que impulse la Transformación Digital de España como una de las palancas fundamentales para relanzar el crecimiento económico, la reducción de la desigualdad, el aumento de la productividad, y el aprovechamiento de todas las oportunidades que brindan estas nuevas tecnologías. Y que lo logre con

respeto a los valores constitucionales y europeos, y la protección de los derechos individuales y colectivos".

5. BIBLIOGRAFÍA

Criado Grande, J. Ignacio y Ramilo Araujo, M. Carmen (2001), "e-administración: ¿un reto o una nueva moda?", en: *Revista Vasca de Administración Pública* 61, nº 1, 11-43.

Liikanen, Erkki (2003), "La administración electrónica para los servicios públicos europeos del futuro", en: *Lección inaugural del curso académico 200-2004 de la UOC*, Barcelona, (en línea), OUC (24/05/04). http://www.uoc./edu/dt/20334/index.html

Lucas Murillo de la cueva, Pablo Lucas (1989/90), "La protección de los datos personales ante el uso de la informática", en: *Anuario de Derecho Público y Estudios Políticos* 2, 153-170.

— (1990) *El derecho a la autodeterminación informativa. La protección de los datos personales ante el uso de la informática*, Tecnos, Madrid.

— (1993) *Informática y protección de datos personales (Estudio sobre la Ley Orgánica 5/1992, de regulación del tratamiento automatizado de los datos de carácter personal)*, Centro de Estudios Constitucionales, Madrid.

— (1999) "La construcción del derecho a la autodeterminación informativa", en: *Revista de Estudios Políticos* 104, abril-junio, 35-60.

Madrid Conesa, Fulgencio (1984) *Derecho a la intimidad, Informática y Estado de Derecho*, Universidad de Valencia, Valencia.

Pérez Luño, Antonio Enrique (1986-87), "La contaminación de las libertades en la sociedad informatizada y las funciones del Defensor del Pueblo", en: *Anuario de Derechos Humanos* 4), 259-289.

— (1987), *Nuevas Tecnologías, Sociedad y Derecho. El impacto socio-jurìdico de las N.T. de la información*, Fundesco, Madrid.

— (1989), "La libertad informática. Nueva frontera de los derechos fundamentales", en: Losano, Mario *et al.* (eds.), *Libertad informática y leyes de protección de datos personales*, Centro de Estudios Constitucionales, Madrid, 185-213.

— (1989/90), "Nuevos derechos fundamentales de la era tecnológica: la libertad informática", en *Anuario de Derecho Público y Estudios Públicos* 2, 171-195.

— (1992), "Del Habeas Corpus al Habeas Data", en *Informática y Derecho* 1, 153-161.

— *Manual de Informática y Derecho*, Ariel, Barcelona, 1996.

— (2000), "Aspectos jurídicos y problemas en Internet", en: De Lorenzo, J. (coord.), *Medios de Comunicación Social y Sociedad: De información a Control y Transformación*, Consejo Social de la Universidad de Valladolid, 107-134.

— *Derechos Humanos, Estado de Derecho y Constitución*, 8ª edic., Tecnos, Madrid, 2003.

Sánchez Bravo, Álvaro (1998), *La protección del derecho a la libertad informática en la Unión Europea*, Publicaciones de la Universidad de Sevilla, Sevilla,1998.

— (1998), *La protección del derecho a la libertad informática en la Unión Europea*, Publicaciones de la Universidad de Sevilla, Sevilla.

— (2001), "Una política comunitaria de seguridad en Internet", en: *Diario La LEY* 5414, 1-8.

— (2010) *A nova sociedade tecnologica: da inclusao ao controle social. A Europ@ é exemplo?,* traduc. de Clovis Gorczevski, EDUNISC, Santa Cruz do Sul.

— (2014) *Derechos humanos y protección de datos personales en el Siglo XXI: homenaje a Cinta Castillo Jiménez*, Punto Rojo Libros, Sevilla.

— (2020) *Derecho, inteligencia artificial y nuevos entornos digitales*, Punto Rojo Libros, Sevilla.

CAPÍTULO XXI

BREVES REFLEXIONES SOBRE LA IMPORTANCIA DEL ESTADO DE DERECHO EN EL DESARROLLO DEL MARCO LEGAL SOBRE LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL EN LA UNIÓN EUROPEA

DIANA CAROLINA WISNER GLUSKO
Centro de Estudios Universitarios "Cardenal Spínola" CEU
Fundación San Pablo Andalucía CEU
cwisner@ceuandalucia.es
https://orcid.org/0000-0003-2723-6112

1. INTRODUCCIÓN

Analizar, estudiar y reflexionar sobre el impacto de la inteligencia artificial (IA) en la sociedad actual es un lugar común en ámbitos multidisciplinares que abordan cada uno de los avances de la llamada cuarta revolución industrial (Parlamento Europeo, 2022, 8), desde diferentes perspectivas.

No existe una única definición de la IA, como tecnología transformadora de nuestro tiempo. El Libro Blanco sobre la inteligencia artificial -un enfoque orientado a la excelencia y la confianza- (en delante, Libro Blanco sobre IA) la define a partir de sus tres elementos esenciales: conjunto de datos, conjunto de algoritmos, y capacidad informática (Comisión, 2020, 3). Sin embargo, otras conceptualizaciones vertidas por la misma institución europea aportan mayor precisión al término que nos ocupa, a través de ejemplos concretos: "Los sistemas basados en la IA pueden consistir simplemente en un programa informático (p. ej. Asistentes de voz, programas de análisis de imágenes, motores de búsqueda, sistemas de reconocimiento facial y de voz), pero la IA también puede estar incorporada en dispositivos de hardware (p. ej. robots avanzados, automóviles autónomos, drones o aplicaciones del internet de las cosas)" (Comisión, 2018, 1).

Es posible distinguir cuatro importantes características que otorgan a la IA el más que suficiente interés para ser objeto de estudio e investigación en el ámbito académico, y que son la transversalidad de las disciplinas involucradas en su diseño, desarrollo, aplicación y control; el hecho de pertenecer a una familia de tecnologías en rápida evolución (Parlamento Europeo y Consejo 2021, 1); el enfoque antropocéntrico respaldado por la Comisión Europea (Comisión Europea, 2019, 2); y la internacionalización de su enfoque regulatorio, en el marco de la implantación de estas tecnologías.

Precisamente el gran potencial que supone la utilización de soluciones basadas en IA, plantea nuevos retos éticos y jurídicos en el entorno mundial, en general, y europeo, en particular; retos que tienen que ver con los principios y valores que sustentan el ordenamiento jurídico y la vida en sociedad.

Como sabemos el Estado de Derecho es un valor común a la UE y a sus Estados Miembros, e implica que la organización política de la comunidad está orientada a la limitación del poder para preservar una esfera autónoma de acción y de realización a los ciudadanos.

Dada la perspectiva internacional de las tecnologías disruptivas en estudio, resulta enriquecedor traer a colación la definición de Naciones Unidas sobre el Estado de Derecho, entendiendo que es "un principio de gobernanza en el que todas las personas, instituciones y entidades, públicas y privadas, incluido el propio Estado, están sometidas a leyes que se promulgan públicamente, se hacen cumplir por igual y se aplican con independencia, además de ser compatibles con las normas y los principios internacionales de derechos humanos" (Consejo de Seguridad de Naciones Unidas, 2004, 5).

Debemos recordar que el Estado de Derecho nació como una fórmula de compromiso que implicaba aunar diversas garantías formales (división de poderes y principio de legalidad, consagrados en la Constitución) con una serie de garantías materiales "ya que el primado de la ley reposaba en su carácter de expresión de la voluntad general y en su inmediata orientación, a la defensa de los derechos y libertades de los ciudadanos" (Pérez Luño, 1986, 220). Es decir que el Estado de Derecho incorpora unos criterios de legitimidad en la organización del poder y de efectividad de los derechos fundamentales. Y, como modelo de Estado, "tiene como objetivo principal la defensa de los derechos de las personas y la eliminación de la arbitrariedad de la actuación de los poderes públicos (López Ulla, 2010, 54).

Sin duda resulta muy interesante, para analizar la importancia del Estado de Derecho en el desarrollo del marco legal sobre los sistemas de IA en la Unión Europea, el planteamiento de Rafael de Asís Roig, explicando el modelo de Peces-Barba[1], quien sostiene que "cuando se utiliza el concepto de Estado de Derecho se usa en dos sentidos diferentes", uno genérico -todo Estado es el hecho fundante básico de un tipo de Derecho, es decir que "todo Estado es Estado de Derecho porque siempre organiza la vida social por medio del Derecho"; y uno específico "que es el liberal, social y democrático de Derecho, que ha incorporado los valores de la ética pública ilustrada, como moralidad política, y que solo con esos rasgos es Estado de Derecho (Asís Roig, 2008, 396).

[1] Peces-Barba, G (1995) *Ética, Poder y Derecho*. Centro de Estudios Constitucionales, Madrid, p. 95.

Teniendo en cuenta esto, el Estado de Derecho (específico) exigiría la adopción de medidas que garantizaran el respeto de los principios de primacía de la ley, la igualdad ante la ley, la separación de poderes, la participación en la adopción de decisiones, la legalidad, la no arbitrariedad, y la transparencia procesal y legal, también frente a la aplicación de soluciones basadas en IA.

El creciente protagonismo de la IA (a nivel técnico, económico, normativo, social, y cultural), hace que esa distinción sea especialmente sensible, dada la profusión de propuestas europeas basadas en una IA confiable y segura, y que plantean el cumplimiento ético y normativo tanto para el sector público como privado, gobernantes y ciudadanos. Qué duda cabe qué hay un gran reto para la cultura jurídica (Pérez Luño, 2021, 44) y que no es otro que dotar a la IA de conciencia[2], en la multiplicidad de ámbitos y sectores en los que actualmente se aplica o se hará en un futuro, ya sea a corto o medio plazo.

Considerando la necesaria interrelación que existe entre el impacto que suponen y supondrán los actuales y futuros desarrollos de la IA -tanto en la UE como en el resto del mundo- y considerando el deber de garantizar uno de los valores y principios esenciales como es el Estado de Derecho, esta contribución se centrará en responder a tres grandes cuestiones. En primer lugar, en qué medida las propuestas europeas para la regulación de la IA se fundamentan en la garantía del Estado de Derecho. En segundo lugar, si el derecho condiciona cómo se desarrollan y se aplican esas tecnologías vinculadas a la IA o si, por el contrario, son estas tecnologías disruptivas las que dan forma al derecho. Y, finalmente, cómo debería ser la inteligencia artificial para no contribuir al debilitamiento del Estado del Derecho e inclusive para mantenerlo y reforzarlo.

2. EL PRINCIPIO DEL ESTADO DE DERECHO EN EL DISEÑO DEL MARCO NORMATIVO DE LA UE SOBRE INTELIGENCIA ARTIFICIAL

Durante el último lustro, han visto la luz diferentes trabajos de las instituciones europeas tendentes a establecer un modelo de gobernanza de la IA y un marco regulador sobre el desarrollo y el impacto de estas tecnologías en la sociedad actual. Todas estas medidas, como no podía ser de otra manera, están orientadas a convertir a Europa en el centro mundial de la inteligencia artificial, con el doble objetivo de "preservar el liderazgo tecnológico de la EU y garantizando que los europeos puedan beneficiarse de las nuevas tecnologías desarrolladas y que funcionan de acuerdo con los valores, los derechos fundamentales y los principios de la Unión" (Parlamento Europeo y Consejo, 2021, 1).

[2] Como explica Pérez Luño, ello implica someter al tribunal de la conciencia, o sea, al conjunto de valores ético-jurídicos, los constantes desarrollos de las nuevas tecnologías y las tecnologías de la información y la comunicación en la experiencia jurídica.

Específicamente, si nos referimos al Estado de Derecho, nadie duda que sea un principio constitucional y estructural común europeo (Weber, 2008, 27) y que, como no podía ser de otra manera, también debe ser reconocido y respetado por las medidas propuestas en materia de IA como uno de los valores en que se fundamenta la Unión.

Para determinar, por tanto, si efectivamente es así, a continuación analizaremos los diferentes textos elaborados por el Parlamento Europeo, el Consejo Europeo y la Comisión Europea sobre IA, en un periodo que abarca desde las primeras comunicaciones en el año 2017, pasando por la Propuesta de Reglamento del Parlamento y del Consejo del pasado año (en adelante, Propuesta de Reglamento de la IA) hasta la reciente Resolución del Parlamento del 3 de mayo de 2022, sobre la inteligencia artificial en la era digital (en adelante, Resolución del Parlamento sobre la IA en la era digital).

Lo primero que llama la atención es que ni la Comunicación de la Comisión relativa a la revisión intermedia de la aplicación de la Estrategia para el Mercado Único Digital, ni la Comunicación de la Comisión Brújula Digital 2030: el enfoque de Europa para el decenio Digital, mencionan expresamente al Estado de Derecho.

Sí lo hace la Comunicación "Inteligencia artificial para Europa", al afirmar que las nuevas tecnologías deben estar basadas en valores porque eso genera confianza en la ciudadanía, lo cual garantizará un marco ético y jurídico adecuado (Comisión, 2018, 16).

Esta misma necesidad de generar confianza en una IA centrada en el ser humano, es la posición adoptada la Comisión Europea en 2019, que no constituye un fin en sí mismo sino un medio al servicio de las personas en pos de su bienestar. Entendiendo que, los valores en los que se basan las sociedades de los Estados miembros -como el Estado de Derecho- "han de estar plenamente integrados en la evolución de la IA" (Comisión, 2019, 2).

En cuanto al Libro Blanco sobre IA -como contribución de la Comisión Europea con propuestas más específicas de acciones de la UE-, a modo introductorio, afirma que generar confianza en la tecnología digital es un requisito previo para su adopción, y ello supone una oportunidad para Europa de liderar mundialmente el desarrollo y la implantación de una IA segura y fiable, dada su estrecha vinculación con los valores y el Estado de Derecho. Y a ello contribuye "su capacidad demostrada de crear productos seguros, fiables y sofisticados en sectores que van desde la aeronáutica a la energía, pasando por la automoción y los equipos médicos (Comisión, 2020, 1). Es decir que, la confianza, como condición anterior y necesaria para el desarrollo de la IA, solo se construye sobre la base de los valores y principios recogidos en los Tratados.

Por su parte el Parlamento Europeo, en su Resolución de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas, menciona expresamente al Estado de Derecho en los considerandos 38 (Responsabilidad social y paridad de género) y 88 (Seguridad de Defensa). En el primero de ellos, destaca que la inteligencia artificial, la robótica y las tecnologías conexas socialmente responsables tienen "un papel que desempeñar en la búsqueda de soluciones que salvaguarden y promuevan los valores fundamentales de nuestra sociedad, como la democracia, el Estado de Derecho, la pluralidad e independencia de los medios de comunicación y una información objetiva y de libre acceso, la salud y la prosperidad económica, la igualdad de oportunidades, los derechos sociales y laborales de los trabajadores, una educación de calidad, la protección de la infancia, la diversidad cultural y lingüística, la paridad de género, la alfabetización digital, la innovación y la creatividad" (Parlamento Europeo, 2020, 14). Por tanto atribuye a la IA la función de adoptar soluciones tendentes a la salvaguarda y la promoción del Estado de Derecho, que no es poco.

En el considerando 88 segundo, resalta que "las políticas de seguridad y defensa de la Unión Europea y de sus Estados miembros se rigen por los principios consagrados en la Carta y por los de la Carta de las Naciones Unidas y por un entendimiento común de los valores universales del respeto de los derechos inviolables e inalienables de la persona, la dignidad humana, la libertad, la democracia, la igualdad y el Estado de Derecho" (Parlamento, 2020, 14). Ello significa que todas las medidas de defensa basadas en IA y adoptadas dentro del marco de la UE deben respetar estos valores universales, fomentando al mismo tiempo la paz, la seguridad y el progreso en Europa y en el mundo.

La Propuesta de Reglamento de la IA, ha supuesto un antes y un después en el diseño del futuro marco regulador en el entorno europeo. Este hito, dentro de las políticas de la UE en el ámbito digital, propone un enfoque basado en riesgos -del todo acertado para combinar la necesaria seguridad jurídica de los proveedores y desarrolladores con la de los usuarios y los afectados- que también resulta insuficiente desde el punto de vista garantista (Cotino Hueso, Salazar, Benjamins *et al.*, 2021).

El texto, que más adelante tendremos la oportunidad de analizar con detenimiento, establece el principio de precaución -en su considerando 15- resaltando que, al margen de los múltiples usos, beneficiosos de la inteligencia artificial, dicha tecnología también puede utilizar indebidamente y proporcionar nuevas y poderosas herramientas para llevar a cabo un cabo prácticas de manipulación, explotación y control social (Parlamento Europeo y

Consejo, 2021, 24). Estas prácticas son sumamente perjudiciales y deben estar prohibidas, pues van en contra de los valores de la Unión de respeto de la dignidad humana, libertad, igualdad, democracia y Estado de Derecho y de los derechos fundamentales que reconoce la UE[3]. Por otro lado, reconoce que deben considerarse de alto riesgo ciertos sistemas de IA aplicados a la administración de justicia y los procesos democráticos, dado que pueden tener potencialmente efectos importantes para la democracia, el Estado de Derecho, las libertades individuales y el derecho a la tutela judicial efectiva ya un juez imparcial.

Finalmente, en 2022, vio la luz la Resolución del Parlamento sobre la IA en la era digital. Un texto extenso en el cual dedica el primer bloque a las posibles oportunidades, riesgos y obstáculos en el uso de la IA -que incluye seis estudios de caso examinados por la Comisión Especial AIDA-, un segundo bloque que analiza el lugar de la EU en la competencia mundial de la IA -desde el aspecto normativo, la situación del mercado y las inversiones- y un tercer bloque donde establece una hoja de ruta para que Europea se convierta en líder mundial en la era digital.

El texto recoge de forma expresa dos alusiones al Estado de Derecho: la primera, relacionada con la política exterior y la dimensión de seguridad de la IA (46) ya que es de imprescindible llegar a un acuerdo mundial sobre normas comunes para el uso responsable de esta tecnología disruptiva, que sea respetuoso con los derechos humanos y el Estado de Derecho. En ese sentido resulta necesaria la cooperación "hacia ciertas normas y principios comunes, normas técnicas y éticas mínimas y orientaciones para un comportamiento responsable a nivel estatal, en concreto a cargo de organizaciones internacionales" (Parlamento, 2022, 19).

Asimismo, pone en valor el hecho de que, aunque en materia de software industrial y robótica la UE va por detrás de Estados Unidos y de China, va por delante en cuanto a estrategias normativas proponiendo un marco normativo horizontal orientado al futuro y favorable a la innovación tanto para el desarrollo, como la implantación y el uso de la IA, con plenas garantías para el Estado de Derecho y los Derechos fundamentales.

Finalizado el análisis, es posible advertir la forma en que ha ido evolucionando el reconocimiento del Estado de Derecho que ha pasado de ser un valor y un elemento que genera confianza en la aplicación de la IA, a ser la piedra angular -junto con los derechos humanos- tanto de los posibles acuerdos internacionales como de la normativa europea que integrará el futuro marco

[3] Como el derecho a la no discriminación, la protección de datos y la privacidad, y los derechos del niño.

regulador de la IA, basado en un sistema de riegos, entre otros, para el mantenimiento del Estado de Derecho, que podrían producirse si se aplicaran determinados sistemas IA[4], sin cumplir con las obligaciones establecidas para cada tipo de riesgo.

3. EL DESARROLLO DE LOS AVANCES TECNOLÓGICOS INTELIGENTES Y LA FORMULACIÓN DEL DERECHO ¿QUIÉN CONDICIONA A QUIÉN?

Uno de los grandes cuestionamientos que se plantean, cuando estamos frente a fenómenos disruptivos o que tienen que ver con la evolución de la técnica, la tecnología o la ciencia, es establecer si realmente estas últimas -de forma directa o indirecta- influyen en cómo será su regulación normativa, o si por el contrario, es el Derecho el que determina cómo se desarrollan y se aplican dichas tecnologías.

Seguramente recordaréis que en los inicios de Internet, existía una corriente de pensamiento en torno a la idea de que regular internet era poner puertas al campo. Lo cierto es que esa corriente supuso un vacío legal en la primera etapa de desarrollo -básicamente por no distinguir claramente la tecnología utilizada y las consecuencias de ese uso u aplicación- y que luego la legislación ha ido corrigiendo.

Si hablamos de IA, Campione nos señala que la hibridación ontológica de lo humano y lo tecnológico está afectando inevitable y radicalmente a la dimensión normativa que ordena nuestra vida (Campione, 2020, 7). Y eso nos lleva a considerar que no solo es la técnica la que condiciona las propuestas sobre futuras normas europeas.

Los modelos de gobernanza de la IA sin duda también influyen en la ecuación que planteamos y en la regulación de la IA. Tenemos por un lado a China y a Estados Unidos, como líderes tecnológicos, con regímenes políticos diferentes, donde los datos personales que son esenciales para el funcionamiento de los algoritmos pertenecen al gobierno -en el primer caso- o a las empresas -en el segundo- ; y por otro lado a la UE, que parte de una regulación cuyo centro es el ser humano, que garantiza el Estado de Derecho y los Derechos Fundamentales, con una legislación armonizada en materia de Protección de Datos Personales frente a sus ciudadanos.

[4] Sistemas de IA prohibidos (implican un riesgo inadmisible para la seguridad, la vida y los derechos fundamentales, por ejemplo, la identificación biométrica). Sistemas de IA de alto riesgo (afectan a los derechos y libertades de las personas aunque si bien no están prohibidos, sí están sujetos a obligaciones reforzadas que garanticen su uso legal, ético, robusto y seguro). Sistemas de IA de riesgo medio/bajo (asistentes virtuales como chatbots). Y sistemas IA de riego mínimo (Filtros de spam, por ejemplo).

Hasta ahora ha sido la ética, es decir las normas deontológicas, las que han ido marcando la senda del futuro desarrollo normativo, tal como hemos tenido oportunidad de apreciar en los epígrafes anteriores.

Precisamente el Grupo de Expertos de Alto Nivel sobre IA publicó los siete requisitos esenciales que deben respetar las aplicaciones de IA para ser fiable, partiendo de la premisa de que, para ser fiable, la misma debe ser conforme a la ley, debe respetar los principios éticos y debe ser sólida, aludiendo con este término a que los algoritmos puedan resolver errores o incoherencias durante todas las fases del ciclo vital (Comisión, 2019, 5). Entre los siete requisitos encontramos la intervención y supervisión humana; la solidez y la seguridad técnica; la privacidad y la gestión de datos; la transparencia; la diversidad, la no discriminación y la equidad; el bienestar social y medioambiental y la rendición de cuentas. Y junto a ellos, no debemos olvidar la regulación de los riesgos de la IA, recogidos en la normativa europea tendente a generar un ecosistema de confianza.

Para responder a la segunda pregunta de las tres planteadas al inicio de esta contribución, es esencial determinar qué debería ser materia de regulación y qué modelo de regulación queremos.

En cuanto a qué regular, debemos distinguir los algoritmos -como un conjunto definido de reglas y procesos para la solución de un problema finito de pasos (Benjamins y Salazar, 2020, 303)-, del código fuente -como base de los programas y de las páginas web sería un bien intangible susceptible de protección como propiedad intelectual o industrial-, y de los robots o máquinas cuyo reconocimiento de personalidad, identidad y responsabilidad continúa siendo objeto de debate, pero que es imposible abarcar en este trabajo, por lo que recomendamos la lectura de los trabajos de Llanos Alonso (2021) y Barrio Andrés (2018).

Tradicionalmente, tanto los algoritmos como los robots eran conceptos que estaban asociados a la técnica, a la tecnología y a la programación. Sin embargo, frente a un cambio de paradigma suscitado a raíz del exponencial desarrollo de la IA, resulta esencial incorporar la perspectiva jurídica, sociológica y filosófica -con el transhumanismo- que aportan una visión multidisciplinar y que analizan los efectos de la aplicación de los algoritmos sobre las personas y sus Derechos Fundamentales, pero también sobre la ventaja competitiva que le otorgan a las empresas e instituciones que la desarrollan. Por tanto, al ser esencial que haya seguridad jurídica, estamos en un Estado de Derecho.

En cuanto a los tipos de estrategias regulatorias, muy brevemente, se habla de hard law (la ley, los reglamentos, el derecho positivo) versus soft law

(protocolos, guías, instrumentos interpretativos, que, sin ser vinculantes ni obligatorios pueden causar efectos jurídicos).

Desde la perspectiva internacional, a nivel Europeo, contamos con la Propuesta de Reglamento de la IA, como futura Ley de Inteligencia Artificial que reviste un alcance general y que será obligatoria en todos sus elementos y directamente aplicable a todos los Estados Miembros. En este sentido, cuando entre en vigor, sus preceptos tendrán legitimidad democrática y fuerza de ley, al contrario que los códigos éticos.

Precisamente la Recomendación de la UNESCO sobre la ética de la inteligencia artificial, recientemente aprobada y que no es legalmente vinculante, pide a sus 193 Estados Miembros que tomen las medidas necesarias para elaborar marcos jurídicos y reguladores a lo largo de todo el ciclo de vida de los sistemas IA, conforme a los principios éticos (UNESCO, 2021, 6).

También es posible plantear si es factible adoptar modelos de corregulación (con la intervención de entes reguladores o de agencias, en materia de certificaciones y estándares) o de autorregulación (como en el sector audiovisual en España). Inclusive la doctrina señala las potenciales utilidades de la ética de la IA pese al natural escepticismo del jurista a través del desarrollo de códigos de conducta y comités ad hoc siendo una pieza más para construir un marco, una estrategia, un sistema y una gobernanza de la IA (Cotino Hueso, 2019, 40).

Entendemos que la autorregulación por sí misma no basta, ni la voluntariedad de quienes deberían cumplir dichas reglas. Por esta razón, la corregulación y la autorregulación deberían considerarse, en relación a la IA, como importantes instrumentos complementarios o suplementarios, pero nunca como alternativa del derecho positivo.

Igualmente es lógico abordar la viabilidad de una regulación general o, por el contrario, sectorial de la IA. Estados Unidos aún no ha introducido legislación horizontal en el ámbito digital, priorizando las leyes sectoriales e incentivando las inversiones y la innovación en IA, aunque poco a poco está comenzando a proporcionar orientaciones jurídicas a las empresas (Parlamento Europeo, 2022, 30).

Lo cierto es que una normativa puramente sectorial y específica dejaría sin regular demasiadas aplicaciones de la IA aplicadas al uso general (Nemitz, 2021, 133) y que contar con una regulación general, serviría de base jurídica y ética para los especiales desarrollos sectoriales en materia de IA, aunque podría haber determinadas áreas sectoriales sin su específica regulación. Esto también tiene que ver con que el planteamiento regulador de la UE se basa el desarrollo de un mercado único digital europeo (Parlamento Europeo, 2022, 31) y en las

consideraciones éticas, en consonancia con los valores fundamentales de los derechos humanos y los principios democráticos.

Por tanto, pensando en un Estado de Derecho, el elegir un tipo de regulación u otro no es una cuestión baladí, pues condicionará al tipo de control que tendrá dicha normativa y las consecuencias de su incumplimiento. Si se legisla a través de un reglamento o de una ley, ante un conflicto, los jueces aplicarían la ley resolviendo conforme a derecho. Si por el contrario se optara por modelos basados en recomendaciones éticas -como la de la UNESCO que prevé un control a través de evaluaciones periódicas de su aplicación- los incumplimientos de los países podrían quedar diluidos y sin consecuencias, dado que carecen de legitimación democrática y de fuerza de ley. Recordemos que esta última resolución aborda la ética de la IA "como una reflexión normativa sistemática, basada en un marco integral, global, multicultural y evolutivo de valores, principios y acciones interdependientes, que puede guiar a las sociedades a la hora de afrontar de manera responsable los efectos conocidos y desconocidos de las tecnologías de la IA en los seres humanos, las sociedades y el medio ambiente y los ecosistemas, y les ofrece una base para aceptar o rechazar las tecnologías de la IA" (UNESCO, 2021, 4).

Por tanto, la UE se encamina y aspira a ser líder tecnológico (algo rezagado) -siguiendo la estela trazada por China y Estados Unidos- y a liderar los procesos de desarrollo de la normativa a través del modelo exportable de una IA de confianza, como organismo mundial de normalización en materia de IA, a través de la coordinación regulatoria y la convergencia de los países socios democráticos. No es tarea fácil establecer los equilibrios necesarios en un Estado de Derecho cuando se opta por regulaciones que incentivan la innovación y la inversión en IA, a costa de una posible merma de las garantías principios y valores que lo sustentan.

Por todo ello, el Parlamento plantea, y pide a la Comisión, que en materia de IA, solo proponga actos legislativos en forma de reglamentos, que sea una legislación flexible, armonizadora, técnicamente neutral, proporcional y basada en el riesgo, respetando los derechos fundamentales (Parlamento Europeo, 2022, 35).

4. HACIA UNA INTELIGENCIA ARTIFICIAL QUE REFUERCE EL ESTADO DE DERECHO O AL MENOS QUE NO CONTRIBUYA A SU DEBILITAMIENTO

En abril de 2018, la Comisión Europea publicó una Estrategia europea para abordar los desafíos y aprovechar las oportunidades que ofrece la IA (Comisión, 2018, 3). Un enfoque en el cual el desarrollo de la IA tiene como centro al ser humano, y se potencia su uso para detectar o curar enfermedades,

anticipar desastres o catástrofes naturales, otorgar mayor seguridad en el transporte, luchar contra la ciberdelincuencia e inclusive mejorar los tiempos de respuesta frente a los ciberataques.

Tanto el Libro Blanco sobre IA como la Propuesta de Reglamento de la IA en la era digital reconocen una serie de beneficios económicos y sociales en todo el espectro de industrias y actividades sociales, y de externalidades positivas por la aplicación de sistemas basados en IA. "Dicha acción es especialmente necesaria en sectores de alto impacto, incluidos el cambio climático, el medio ambiente y la salud, el sector público, las finanzas, la movilidad, los asuntos de interior y la agricultura" (Parlamento Europeo y Consejo, 2021, 1).

Pero como si se tratara de una moneda, con su cara y cruz, también las instituciones europeas señalan de forma indirecta -el Libro Blanco sobre IA- y de forma expresa -la Propuesta de Reglamento de la IA en la era digital- la conformación de diferentes niveles de riesgo, en tanto y en cuanto conculquen derechos fundamentales o menoscaben el Estado de Derecho. En esa dirección Llano Alonso reconoce que "del mismo modo del mismo modo que la IA, la robótica y las tecnologías conexas poseen un enorme potencial para generar oportunidades para empresas y beneficios para los ciudadanos, también suponen un impacto directo sobre nuestros derechos y libertades" (Llano Alonso, 2021, 320).

Por todo ello, la tercera y última de las preguntas que ha guiado este trabajo es del todo oportuna, ya que resulta esencial conocer de qué forma y bajo qué circunstancias los sistemas basados en IA deben ser regulados de manera tal que contribuyan a reforzar el Estado de Derecho o al menos que no lo debiliten.

Según Paul Nemitz, esto último dependerá de las formas de regulación elegidas (Nemitz, 2021, 130), es decir del tipo de regulación por el que se opte, tal como hemos abordado en el epígrafe anterior. Sin embargo, la respuesta al interrogante planteado lo debemos hacer no solo a través del continente (modelo regulatorio) sino del contenido (materia regulada) para conocer, con mayor detenimiento, en qué aspectos el Estado de Derecho podría verse beneficiado o perjudicado ante la aplicación de soluciones basadas en IA.

Los expertos del Comité ad hoc sobre inteligencia artificial del Consejo de Europa, han establecido una serie de premisas que recogen las medidas necesarias para el desarrollo de un marco legal sobre los sistemas de inteligencia artificial, teniendo en consideración los derechos humanos, la democracia y el Estado de Derecho (CAHAI, 2019, 31). Teniendo en cuenta dicho informe, podemos señalar una serie de beneficios y de riesgos generados a partir de la utilización de la IA por determinados sujetos:

1.- La utilización por parte de instituciones y Administraciones Públicas como detentadoras de poder público podría aumentar la eficiencia en sus procedimientos administrativos; aunque existe un alto riego de conculcar el principio de Estado de Derecho si su utilización no estuviera justificada, si no fuese proporcional o si discriminara. Con lo que se produciría una merma de confianza y una pérdida del poder de autoridad que le otorga la ley, frente a los ciudadanos.

Resulta de suma utilidad en este campo, la propuesta de Alejandro Huergo Lora en torno a la importancia de detectar cuál es la función que cumplen los algoritmos en el proceso de creación o aplicación del derecho. En ese sentido distingue tres tipos de algoritmos: los que traducen un régimen jurídico para facilitar la toma de decisiones por la Administración (cálculo de un tributo), los que sirven para mecanizar o automatizar procesos reglados, sin cambiar su marco normativo, pero no se puede prescindir de él al momento de controlar la actuación administrativa (casos de asignación de recursos escasos) y, finalmente, los de tipo predictivo que contribuyen a orientar en una determinada dirección la actuación administrativa y que, a diferencia de los anteriores, aportan elementos decisionales propios (Huergo Lora, 2021, 1).

En esa dirección de pensamiento, creemos que resulta fundamental tener en cuenta el impacto teórico, en cuanto a beneficios que supone la aplicación de la IA frente a los efectos reales de su aplicación como se ha visto en algunos casos de discriminación. Basta mencionar la Sentencia de 5 de febrero de 2020 del Tribunal Distrito de La Haya -de Primera Instancia- [5]en la cual reconoce que un sistema de análisis denominado SyRI (System Risk Indication), usado por el Gobierno, para rastrear posibles fraudes al Estado (residentes de determinados barrios problemáticos, o en situación de pobreza), no respetaba la privacidad del ciudadano, vulnerando el artículo 8 de la Carta de los Derechos Fundamentales de la Unión Europea. Por tanto, es necesario que el marco regulador establezca la obligación de los gobiernos de detener activamente las aplicaciones que pudieran aumentar la desigualdad entre los ciudadanos.

2.- En cuanto a los Tribunales y las fuerzas del orden y Cuerpos de Seguridad del Estado, estos podrían volverse más eficientes en la lucha contra la delincuencia, el terrorismo, el crimen organizado, si utilizaran determinados sistemas biométricos de vigilancia. Esto sería posible, aunque con el riesgo de volverse más opaca y menos auditable, lo que dificultaría o imposibilitaría

[5] Rechtbank Den Haag, 05-02-2020, C-09-550982-HA ZA 18-388. ECLI:NL:RBDHA:2020:865 https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBDHA:2020:1878&showbutton=true&keyword=syri

saber si existe un impacto real en los derechos humanos, la democracia y el estado de derecho

3.- En cuanto al uso de soluciones basadas en IA para la eficiencia de la Justicia, ya en 2018 la Comisión Europea estableció cinco 5 principios para el uso de la IA en el poder judicial en la "Carta Ética Europea sobre el uso de la IA en los sistemas judiciales y su entorno", entre ellos podemos mencionar el respeto de los derechos fundamentales (tanto en el diseño como la implementación de herramientas); el principio de no discriminación; el principio de calidad y seguridad con respecto al procesamiento de decisiones judiciales; la transparencia, imparcialidad y equidad y el principio "bajo el control del usuario" centrado en garantizar el deber de información (Comisión Europea, 2018b, 4 y siguientes).

4.- El desarrollo y el despliegue por parte de los grandes grupos tecnológicos, por un lado, ayuda a la aceleración de la industria vinculada a esta tecnología disruptiva pero, por el otro, al ser capaces de controlar la totalidad del ecosistema donde se implantan soluciones de IA, podrían determinar y quizás alterar las estructuras sociales e incluso democráticas, con un gran impacto en los derechos humanos.

Son muchos de estos sistemas privados de IA determinan qué manifestaciones de los usuarios deben eliminarse o no de las redes sociales según detecten mediante algoritmos los llamados "discursos del odio". Anteriormente eran los Tribunales los únicos que determinaban el alcance de la libertad de expresión. Hoy, estos sistemas de IA compiten, de facto, por la autoridad con los jueces y la ley, desarrollándose "sistemas" que operan fuera de los límites y las protecciones que brinda el Estado de Derecho.

Otro tanto sucede con las resoluciones automatizadas de solución de controversias online, proporcionadas por empresas privadas, se rigen por una legislación que podría mermar la protección que brindan los derechos de los consumidores y usuarios, tanto en lo que se refiere a las vías de reclamación judicial como extrajudicial.

5.- En lo que respecta la utilización de la IA por parte del poder legislativo, Anthony Casey y Anthony Niblett, señalan que la utilización de una tecnología predictiva, impulsada por una capacidad computacional cada vez mayor, permitirá a los legisladores esculpir leyes ex ante cada vez más perfectas (Casey y Niblett, 2017, 1410). De esa manera, las leyes estarán altamente calibradas de acuerdo con los objetivos políticos, lo que quitaría la posibilidad a futuras interpretaciones por parte de los jueces. Y no solo eso, sino que confiando en las máquinas para observar y analizar hechos más relevantes, los

legisladores harían mejores predicciones sobre el impacto de una ley y por tanto se cometerían menos errores al momento de legislar.

Dicho esto, el Grupo de Expertos CAHAI va más allá y señala una serie de elementos que contribuirían al fortalecimiento del Estado de Derecho. Tal es el caso de los desarrollos algorítmicos que facilitan a los reguladores o agencias a identificar los casos de falseamiento de la competencia, corrupción; de igual forma que puede utilizarse para detectar y defenderse contra ataques cibernéticos (CAHAI, 2019, 34). En ese sentido, resulta del todo interesante, la aportación conjunta de la Comisión Nacional de los Mercados y la Competencia (CNMC) y la Autoritat Catalana de la Competència (ACCO) informando sobre la necesidad de una adaptación de la normativa de tal manera que las autoridades de competencia puedan también hacer uso de la IA, así como que estos organismos públicos puedan ser más permeables al conocimiento existente en este ámbito del conocimiento (Comisión Nacional de los Mercados y la Competencia (CNMC) y la Autoritat Catalana de la Competència, 2020, 1).

Finalmente, teniendo en cuenta los siete principios del Grupo de Expertos de Alto Nivel sobre IA que deben respetar las aplicaciones para que sea fiable, la Propuesta de Reglamento de la IA y la Resolución del Parlamento Europeo sobre la IA en la era digital, las líneas directrices aplicables para que el desarrollo y la utilización de la IA refuerce o al menos no contribuya al debilitamiento del Estado de Derecho serían las siguientes:

En primer lugar, que el ser humano y su inalienable dignidad conserven la centralidad en el desarrollo normativo y en el uso de sistemas basados en IA, tanto en el ámbito público como privado. Porque como sostiene Suñé Llinás, "el ser humano en cuanto especie puede verse desplazado en su centralidad por la tecnología y no sólo por la biotecnología -especialmente la ingeniería genética- sino también por la informática y muy señaladamente la inteligencia artificial" (Suñé Linás, 2021, 213).

En segundo lugar, debemos situar los derechos humanos como piedra angular del desarrollo de la IA, lo cual se traduce en la necesaria garantía del derecho a la no discriminación cualquiera que fuera su origen, causa o naturaleza, en relación con las decisiones, uso de datos y procesos basados en IA. Existe un gran desconocimiento de los impactos potenciales de la IA en la garantía de los derechos humanos, por lo que deviene esencial crear verdadera consciencia de ello en la ciudadanía, en las empresas y en el sector público.

En tercer término, que el cumplimiento de las obligaciones, rendición de cuentas y reparación del daño, tienen que ver con el principio de legalidad y, fundamentalmente, con una regulación basada en el riesgo que supone la utilización de aplicaciones de IA. Y ello deriva en que, ante situaciones que

pudieran producir daños en los bienes y derechos de las personas -en su esfera personal y patrimonial-, estas deben contar con procedimientos judiciales y extrajudiciales para impugnar decisiones basadas en algoritmos.

En cuarto lugar, resaltar que la supervisión humana ayuda a garantizar que un sistema de IA no socave la autonomía humana o causea otros efectos adversos. Esta supervisión requiere un análisis del impacto económico, social, legal y ético de la decisión basada en IA. El solo hecho de contar con la supervisión humana, posibilita decidir la viabilidad de la decisión tomada, incluyendo la capacidad de anularla. También a través de la supervisión humana se podría prevenir la influencia electoral o la manipulación pública, en definitiva, proteger la democracia, las estructuras democráticas, la libertad de expresión y el Estado de Derecho.

En quinto lugar, la explicabilidad, la auditabilidad, la trazabilidad, la transparencia y la motivación son garantías esenciales y necesarias para las decisiones que se tomen utilizando algoritmos. Todas ellas constituyen formas de rendición de cuentas que contribuyen al sostenimiento del Estado de Derecho.

En sexto lugar, es esencial que se respete la privacidad de los datos en todo el ciclo de vida de los sistemas basados en IA, con las suficientes garantías de que se realizará un correcto tratamiento de los datos, conforme a la normativa vigente.

Y, por último, no debe perderse de vista la perspectiva social de la IA, es decir el impacto a corto, medio y largo plazo de la aplicación de sus sistemas y las medidas adoptadas para prevenir o mitigar los daños, algo que sin duda reforzará el Estado de Derecho en tanto y en cuanto estos sistemas se apliquen para mejorar la vida en sociedad y ser facilitadores del bienestar social general.

5. CONCLUSIONES

Siguiendo el hilo reflexivo planteado desde el inicio del presente capítulo, podemos afirmar que los trabajos, propuestas y contribuciones de las instituciones europeas analizadas reconocen la importancia del Estado de Derecho desde la misma base que da fundamento al planteamiento de un marco regulador europeo de la IA. Y no solo hablamos del hecho que la expresión "Estado de Derecho" se encuentre expresamente recogida, sino de algo de mayor calado, como es que los aspectos regulatorios deberán siempre ser diseñados, desarrollados, y aplicados a la luz de los principios y valores que sustentan la EU. Sin lugar a duda los modelos de gobernanza de la IA son decisivos para optar por sistemas más o menos garantistas de los derechos humanos. La competencia entre los tres grandes bloques, que una y otra vez se recogen en los textos analizados (China, Estados Unidos y la UE) no deja de ser

una competencia de valores, donde la centralidad del ser humano es la clave que marca la diferencia.

En la historia del desarrollo de la técnica, la ciencia y la tecnología, siempre ha habido un punto de condicionamiento de dichos avances en relación a su regulación jurídica. En el caso de la IA, es claro que las consecuencias de la aplicación de sistemas basados en IA -como los casos comprobados y detectados de discriminación algorítmica- contribuyen a plantear soluciones normativas que refuercen la equidad y la no discriminación. Por ello, ante la aparición de nuevas tecnologías disruptivas es la regulación legal y no la desregulación -al menos en una primera etapa- lo que beneficia al interés público, sin olvidar que la ética da forma o moldea a la ley.

Hemos podido comprobar cómo, desde la UE, se plantea la exportación de un modelo de IA "fiable", recurriendo al reglamento europeo como acto legislativo para otorgar la suficiente fuerza al marco normativo garantista que se plantea basado en un sistema de riesgos. Y este planteamiento se realiza aún siendo conscientes de que Estados Unidos y China, son tecnológicamente más potentes y más avanzados en esta materia. Por tanto, sin duda alguna el condicionamiento entre la tecnología y el derecho podría ser recíproco, si lo analizamos desde una doble perspectiva tanto temporal como espacial.

La aplicación de determinados algoritmos sin ninguna duda nos plantea nuevos desafíos regulatorios, sociales y éticos, en relación a las empresas que los desarrollan e implementan como a las Administraciones Públicas, el legislador o la Justicia que los aplican. En cuanto a la finalidad en el uso de algoritmos, el punto de partida sería propiciar el bienestar social, el desarrollo tecnológico y el reconocimiento de la dignidad de las personas y del resto de derechos fundamentales, en un Estado de Derecho. Por tanto, en la medida en que la regulación algorítmica se convierta en parte de las prácticas legislativas, judiciales o administrativas, debemos asegurarnos de que no sea simplemente compatible con el Estado de Derecho, sino que, efectivamente y de forma concreta, integre sus principios y presupuestos básicos.

Si efectivamente se cumpliera y se cristalizara en un marco normativo de obligado cumplimiento, existirían más posibilidades de que no se produjeran sesgos ni discriminación en la aplicación de soluciones basadas en IA; o que si se produjeran, el mismo sistema -conforme al principio de supervisión humana y de acuerdo a la legitimidad que le otorgan las normas- aplicaría los protocolos asociados a la explicabilidad, la auditabilidad, la rendición de cuentas y la responsabilidad, junto a la posibilidad de impugnación de los usuarios, en cumplimiento de la legislación vigente.

En definitiva, que el ser humano y su inalienable dignidad conserven la centralidad en el desarrollo normativo y en el uso de sistemas basados en IA desde la perspectiva europea, nos demuestra el grado de importancia que posee el Estado de Derecho en el desarrollo del marco legal sobre los sistemas de inteligencia artificial en la UE.

6. BIBLIOGRAFÍA

Barrio Andrés, Moisés (2018), *Derecho de los robots*, Madrid, Wolters Kluwer.

Benjamins, Richard, Idoia Salazar (2020), *El mito del algoritmo. Cuentos y cuentas de la Inteligencia artificial*, Anaya, Madrid.

CAHAI Secretariat (2020), *Global perspectives on the development of a legal framework on Artificial Intelligence systems based on the Council of Europe's standards on human rights, democracy and the rule of law*. Compilations of contributions DGI (2020) 16.

Casey, Anthony J., Anthony Niblett (2017), "La muerte de las reglas y los estándares", en *Indiana Law Journal*, 92, edición 4, artículo 3, 1401-1447. https://www.repository.law.indiana.edu/ilj/vol92/iss4/3

Carta de los Derechos Fundamentales de la Unión Europea (2000). DO C. 364. 18.12.2000. https://www.europarl.europa.eu/charter/pdf/text_es.pdf

Comisión Europea (2017). Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones relativa a la revisión intermedia de la aplicación de la Estrategia para el Mercado Único Digital Un mercado único digital conectado para todos. COM/2017/0228 final. https://eur-lex.europa.eu/resource.html?uri=cellar:a4215207-362b-11e7-a08e-01aa75ed71a1.0005.02/DOC_1&format=PDF

— (2018a). Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. Inteligencia artificial para Europa. COM/2018/237 final. https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52018DC0237&from=ES

— (2018b) "Carta Ética Europea sobre el uso de la IA en los sistemas judiciales y su entorno". *CEPEJ* (2018) 14. 3.12.2018.

— (2019) Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de regiones. Generar confianza en la inteligencia artificial centrada en el ser humano. COM (2019) 168. 08.04.2019. https://eur-lex.europa.eu/legal-content/ES/ALL/?uri=CELEX:52019DC0168

— (2020). Libro Blanco sobre inteligencia artificial - Un enfoque europeo orientado a la excelencia y la confianza. COM (2020) 65 final, 19.02.2020. https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_es.pdf

— (2021). Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. Brújula Digital 2030: el enfoque de Europa para el decenio Digital. COM (2021) 118 final. 9.03.2021. https://eur-lex.europa.eu/resource.html?uri=cellar:12e835e2-81af-11eb-9ac9-01aa75ed71a1.0022.02/DOC_1&format=PDF

Comisión Nacional de los Mercados y la competencia y la Autoritat Catalana de la Competència (2020). Inteligencia y Competencia. https://www.cnmc.es/sites/default/files/editor_contenidos/Notas%20de%20prensa/2020/CONTRIBUCI%C3%93N%20IA%20Y%20COMPETENCIA%20CNMC%20ACCO.pdf

Consejo de Seguridad de Naciones Unidas (2004). El Estado de Derecho y la justicia de transición en las Sociedades que sufren o han sufrido conflictos. Informes del Secretario general. 3 de agosto de 2004. https://documents-dds-ny.un.org/doc/UNDOC/GEN/N04/395/32/PDF/N0439532.pdf?OpenElement

Cotino Hueso, Lorenzo (2019), "Ética en el diseño para el desarrollo de una inteligencia artificial, robótica y big data confiables y su utilidad desde el derecho", en: *Revista Catalana de Dret Públic*, 58, 29-48. https://doi.org/10.2436/rcdp.i58.2019.3303

Cotino Hueso, Lorenzo *et al.* (2021), "Un análisis crítico constructivo de la Propuesta de Reglamento de la Unión Europea por el que se establecen normas armonizadas sobre la Inteligencia Artificial (Artificial Intelligence Act)", en: *Diario La Ley*, 2 de julio de 2021, Wolters Kluwer.

España (2018). Ley orgánica 3/2018, de 5 de diciembre, de protección de Datos Personales y garantía de los derechos digitales. Boletín Oficial del Estado, 6 de diciembre de 2018, núm. 294.

Huergo Lora, Alejandro (2021), Regular la inteligencia artificial (en Derecho Administrativo), en: *El Blog. Revista de Derecho Público*. 8 de marzo. http://blogrdp.revistasmarcialpons.es/blog/regular-la-inteligencia-artificial-en-derecho-administrativo-por-alejandro-huergo-lora/

Llano Alonso, Fernando (2021), "De máquinas y hombres. Tres cuestiones ético-jurídicas sobre la Inteligencia Artificial", en Llano Alonso, Fernando, Garrido Martín, Joaquín (eds.), *Inteligencia artificial y Derecho. El jurista ante los retos de la era digital*, Aranzadi, Navarra, 201-234.

— (2022). "El jurista ante los retos de la revolución tecnológica 4.0", en Sánchez Bravo, Álvaro (ed.), *Semper Sapiens: Libro homenaje al Prof. Felipe Rotondo Tornaría,* Alma Mater, Madrid, 305-324.

López Ulla, Juan Manuel (2010). "Defensa de la Constitución: Jurisdicción Constitucional, Reforma y Estados excepcionales", en Agudo Zamora, Miguel *et al.* (eds.), *Manual de Derecho Constitucional*, Tecnos, Madrid, 53-83.

Nemitz, Paul (2021), "La democracia en la era de la inteligencia artificial", en: *Revista Nueva Sociedad,* 294, de julio-agosto de 2021, 130- 140.

Parlamento Europeo (2020). Resolución del Parlamento Europeo, de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas (2020/2012 (INL)). 20.10.2020. https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_ES.pdf

— (2021). Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia artificial) y se modifican determinados actos legislativos de la Unión. COM (2021) 206 final, 21.04.2021. https://eur-lex.europa.eu/legal-content/EN/TXT/?qid= 1623335154975&uri=CELEX%3A52021PC0206

— (2022). Resolución del Parlamento Europeo, de 3 de mayo de 2022, sobre la inteligencia artificial en la era digital (2020/2266 (INI)). P9_TA (2022)0140. https://www.europarl.europa.eu/doceo/document/TA-9-2022-0140_ES.pdf

Pérez Luño, Antonio (1986), *Derechos Humanos, Estado de Derecho y Constitución*, Tecnos, Madrid.

— (2021), "La inteligencia artificial en tiempos de pandemia", en Llano Alonso, Fernando, Garrido Martín, Joaquín (eds.), *Inteligencia artificial y Derecho. El jurista ante los retos de la era digital,* Aranzadi, Navarra, 33-50.

Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE. Diario Oficial de la Unión Europea L nº 119, de 4 de mayo de 2016.

Suñé Llinás, Emilio (2022), *Derecho e Inteligencia Artificial. De la robótica a lo posthumano*, Tirant Lo Blanch, Ciudad de México.

Tratado de Funcionamiento de la Unión Europea. Diario Oficial de la Unión Europea. C 83/47, 30.03.2010. https://www.boe.es/doue/2010/083/Z00047-00199.pdf

Tratado de la Unión Europea. Diario Oficial de la Unión Europea. C 83/13. 30.03.2010. https://www.boe.es/doue/2010/083/Z00013-00046.pdf

UNESCO (2021). Recomendación sobre ética de la Inteligencia Artificial. https://unesdoc.unesco.org/ark:/48223/pf0000380455_spa

Weber, Albrech (2008). "El principio de Estado de Derecho como principio constitucional común europeo", en *Revista Española de Derecho Constitucional*, 84, septiembre-diciembre, 27-59.

SOBRE LOS AUTORES

RAFAEL DE ASÍS ROIG

Catedrático de Filosofía del Derecho de la Universidad Carlos III de Madrid. Co-director del Grupo de Investigación Derechos humanos, Estado de Derecho y Democracia, Responsable de la Red "El tiempo de los derechos" y de la Clínica Javier Romañach. Autor de más de un centenar de publicaciones, entre las que cabe destacar los siguientes trabajos directamente relacionados con la temática de este libro: *Una mirada a la robótica desde los derechos humanos*, Dykinson, Madrid, 2014; "Robótica, Inteligencia Artificial y Derecho", *Revista de privacidad y Derecho digital*, vol. 3, nº 10, abril-junio, 2018, pp. Pp. 27-77; "Desafíos éticos de los ciborgs", *Universitas: Revista de Filosofía, Derecho y Política*, nº 30, 2019, pp. 1-25; *Derechos y tecnologías*, Dykinson, Madrid, 2022; "Sobre la propuesta de los neuroderechos", *Derechos y libertades. Revista de Filosofía del Derecho y Derechos Humanos*, nº 47, 2022, pp. 51-70.

NURIA BELLOSO MARTÍN

Catedrática de Filosofía del Derecho en el Departamento de Derecho Público de la Facultad de Derecho de la Universidad de Burgos. Desde noviembre de 2011 hasta finales de 2022 fue Directora del Departamento de Derecho Público. Entre 1996 y 2016 fue Coordinadora del Programa de Doctorado "Sociedad plural y nuevos retos del Derecho", del Departamento de Derecho Público. Directora del Curso de Especialista en Mediación Familiar de la UBU, desde el año 2002. Es miembro del Consejo editorial internacional de la revista científica *Ethikai* y del *EthicAI Institute -Ethics in Artificial Intelligence*. Ha publicado numerosos trabajos sobre protección de derechos fundamentales en la era digital (2009), aplicación de nuevas tecnologías a centros penitenciarios (2015), Inteligencia Artificial y algunas de sus aplicaciones en el ámbito legislativo y en el ámbito judicial (2020), y desafíos iusfilosóficos de los usos de la inteligencia artificial en los sistemas judiciales (2021).

STEFANO BINI

Profesor Ayudante Doctor acreditado Profesor Contratado Doctor de Derecho del Trabajo y de la Seguridad Social en la Universidad de Córdoba. Es Doctor cum laude en Derecho por la Universidad de Sevilla, y en Derecho y Empresa por la Universidad LUISS Guido Carli de Roma. Como investigador está especializado en el impacto de la digitalización en la dimensión individual y colectiva de las relaciones laborales. Entre sus publicaciones más recientes destaca su monografía titulada: *La dimensión colectiva de la digitalización del trabajo*, Bomarzo, Albacete, 2021.

ROGER CAMPIONE

Profesor Titular de Filosofía del Derecho de la Universidad de Oviedo (acreditado a catedrático). Doctor Europeo por la Universidad de Oviedo y *Dottore in Giurisprudenza* por la Universidad de Pisa. Académico correspondiente de la Real Academia Asturiana de Jurisprudencia. En la actualidad dirige el Proyecto de I+D+i del Plan Nacional Retos de la Sociedad "El logos de la guerra. Normas y problemas de los conflictos armados actuales". Es autor de diversas monografías y decenas de publicaciones en el ámbito de la filosofía jurídica y la teoría social, entre las cuales cabe destacar su monografía *La plausibilidad del Derecho en la era de la inteligencia artificial. Filosofía carbónica y filosofía silícica del Derecho*, Dykinson, Madrid, 2020.

THOMAS CASADEI

Catedrático de Filosofía del Derecho en la Facultad de Derecho de la Universidad de Modena-Reggio Emilia (Italia). Ha sido responsable de LABdi - Laboratorio su forme di discriminazione, istituzioni, azioni positive (2007-2010), Vicedecano de la Facultad de Derecho de la Universidad de Modena-Reggio Emilia (2018-2019) y, desde noviembre de 2019, Delegado de Comunicación y Portavoz del Rector de la Universidad de Modena-Reggio Emilia. Fue uno de los fundadores del CRID - Centro di Ricerca Interdipartimentale su Discriminazioni e vulnerabilità, a cuya junta directiva pertenece desde 2016. Codirige, junto al Prof. Zanetti, las colecciones del Centro "Diritto e vulnerabilità" (Giappichelli) y "Prassi sociale e teoria giuridica" (Mucchi). Es también codirector de otras colecciones jurídicas, como "Comp.lex. Diritto, computazione, complessità" (Wolters Kluwer), junto a Stefano Pietropaoli, y "Altera pars. Studi e ricerche di Filosofia politica e Teoria del diritto" (Liberedizioni), junto a Roberto Cammarata. Autor de más de un centenar de publicaciones, en su bibliografía destacan tres manuales: *La didattica del diritto*, Pacini Editore, Pisa, 2019; 2021 (2ª ed.); escrito junto a V. Marzocco y S. Zullo; *Manuale di Filosofia del Diritto. Figure, categorie, contesti*, Giappichelli, Torino, 2019; 2020 (2ª ed.); escrito junto a Gf. Zanetti); *Diritto e tecnologie informatiche. Questioni di informatica giuridica, prospettive istituzionali e sfide sociali*, Wolters-Kluwer-Cedam, Milano, 2021; coeditado con Stefano Pietropaoli.

MIGUEL DE ASÍS PULIDO

Investigador y Becario FPI en la Facultad de Derecho de la UNED, Programa de Doctorado en Derecho y Ciencias Sociales. Graduado en Derecho y Administración y Dirección de Empresas por la Universidad Carlos III de Madrid en 2019. Máster de Acceso a la Abogacía por la UNED en 2021. Ha trabajado como becario en el Ministerio de Justicia y en el Banco de España.

Actualmente se encuentra realizando su tesis doctoral sobre la influencia de la Inteligencia Artificial en el derecho al proceso debido, sobre el que ya ha publicado un capítulo en el libro *Inteligencia Artificial y Derecho. El jurista ante los retos de la era digital* (coeditores: Fernando H. Llano Alonso y Joaquín Garrido Martín), Thomson Reuters Aranzadi, Cizur Menor (Navarra), 2021.

DANIEL GARCÍA SAN JOSÉ

Catedrático de Derecho Internacional Público en la Facultad de Derecho de la Universidad de Sevilla, donde es Vicedecano de Relaciones Internacionales y Movilidad desde 2014 hasta la actualidad. Diploma del Centre d'Etudes et des Recherches de l'Académie de Droit International de La Haye y Diploma with distinction del Erik Castren Institut de Derecho Internacional de Helsinki. Investigador honorario de la Universidad Autónoma de Chile. Fundador y Co-Director de la revista semestral *Ius et Scientia. Revista Electrónica de Derecho y Ciencia* Es autor de numerosas publicaciones científicas en las áreas de Docencia Universitaria, Derechos Humanos, Medio Ambiente, Bioderecho y Derecho Interestelar. En su amplia bibliografía destacan dos monografías recientes: *Interstellar Law: Ius Gentium for New Worlds*, Laborum, Murcia, 2018; *La libertad de expresión 4.0 en el sistema del Convenio europeo de derechos humanos*, Tirant lo Blanch, Valencia, 2022.

JOAQUÍN GARRIDO MARTÍN

Profesor Ayudante Doctor de Derecho Romano en la Facultad de Derecho de la Universidad de Sevilla. Licenciado en Derecho (2012) y en Filosofía (2015) por la Universidad de Sevilla. Ha realizado estancias de investigación en las universidades de Berna, Instituto Universitario Europeo de Florencia, Oxford, Múnich, Frankfurt, Hamburgo y Heidelberg, donde también imparte cursos de grado y posgrado. Recientemente ha ganado la prestigiosa Beca Humboldt como investigador en el Instituto Max Planck de Hamburgo. Entre sus publicaciones más recientes, relacionadas con la temática de la presente obra, destaca el capítulo: "Inteligencia Artificial y cultura tecnológica. Hacia una técnica fragmentada", en el volumen colectivo *Inteligencia Artificial y Derecho. El jurista ante los retos de la era digital* (coeditores: Fernando H. Llano Alonso y Joaquín Garrido Martín), Thomson Reuters Aranzadi, Cizur Menor (Navarra), 2021.

ANA GARRIGA DOMÍNGUEZ

Profesora Titular de Filosofía del Derecho en la Facultad de Derecho de la Universidad de Vigo. Ha impartido o imparte docencia de posgrado en diversos másteres: Master sobre Derechos Fundamentales de la Universidad Carlos III de Madrid; Master en Consultoría de Software Libre de la Escuela

Superior de Ingeniería Informática de la Universidad de Vigo, Master de Ordenación Jurídica del Mercado de la Universidad de Vigo; Maestría en Derechos Humanos y Democratización en la Universidad Externado de Colombia; Master en Derecho Urbanístico y Medio Ambiente de la Universidad de Vigo, así como en varios cursos de especialista y doctorado. Ha sido investigadora responsable de varios proyectos de I+D+i financiados tanto por el Ministerio de Educación y Ciencia, como por la Xunta de Galicia. Miembro del Consejo de redacción de la revista *Derechos y libertades*. Especialista en protección de datos personales e Informática jurídica, entre sus publicaciones destacan: *Nuevos retos para la protección de datos personales: en la era del big data y de la computación ubicua*, Dykinson, Madrid, 2015; *Un nuevo reto para los derechos fundamentales: los datos genéticos* (codirectoras: Ana Garriga Domínguez y Susana Álvarez González), Dykinson, Madrid, 2017.

LAURA GÓMEZ ABEJA

Profesora Ayudante Doctora de Derecho Constitucional en la Facultad de Derecho de la Universidad de Sevilla. Es doctora en Derecho Constitucional por esta misma Universidad. Su tesis doctoral, *Las objeciones de conciencia (Centro de Estudios Políticos y Constitucionales, Madrid*, 2016), obtuvo un sobresaliente *cum laude* por unanimidad y recibió una mención especial del Jurado del Premio Nicolás Pérez Serrano, convocado por del Centro de Estudios Políticos y Constitucionales (2013). Ha participado en el Proyecto de Investigación de I+D+i ; actualmente es miembro del equipo de investigación del Proyecto de I+D+i "Biomedicina, Inteligencia Artificial, Robótica y Derecho: los Retos del Jurista en la Era Digital" (PID2019-108155RB-I00) Entre sus publicaciones destacan sus monografías *Derecho a rechazar el tratamiento médico: análisis de los antecedentes desde una perspectiva constitucional,* Tirant Lo Blanch, Valencia, 2014, así como diversos artículos sobre la objeción de conciencia, la obediencia al Derecho o el consentimiento informado.

Mª ISABEL GONZÁLEZ TAPIA

Profesora Titular de Derecho Penal en la Facultad de Derecho de la Universidad de Córdoba y abogada. Ha realizado estancias de investigación en las universidades de Cambridge y Edimburgo. Una de sus principales líneas de investigación y de especialización es el neuroderecho, la neurobiología del comportamiento antisocial y sus implicaciones en el Derecho penal, ámbito en el que ha dirigido un Proyecto de I+D+i financiado por el Ministerio de Economía y Competitividad. Entre sus publicaciones relacionadas con el perfil temático de este volumen destaca sus artículos: "Bad Genes y responsabilidad criminal", *Revista de derecho y genoma humano: genética, biotecnología y medicina avanzada*, nº 1, 2014, pp. 313-317 y "A New Legal Treatment for Psychopaths?

Perplexities for Legal Thinkers" (en coautoría con Ingrid Obsuth y Rachel Heeds), *International Journal of Law and Psychiatry*, 2017, open access.

FERNANDO H. LLANO ALONSO

Catedrático de Filosofía del Derecho de la Facultad de Derecho de la Universidad de Sevilla, donde es Vicedecano de Investigación y Doctorado desde 2014 hasta la actualidad. Ha realizado estancias de investigación en las universidades de Bolonia, Pavía, Trieste y Pisa (Italia), Coimbra (Portugal), Maguncia (Alemania), Edimburgo y Oxford (en donde ha sido *Academic Visitor* entre los años 2010 y 2019). Investigador responsable del Proyecto de I+D+i del Ministerio de Ciencia e Innovación "Biomedicina, Inteligencia Artificial, Robótica y Derecho: Los retos del jurista en la era digital" (PID2019-108155RB-I00). Es responsable del grupo de investigación SEJ504: "Bioderecho Internacional". Co-Director de la revista semestral *Ius et Scientia. Revista Electrónica de Derecho y Ciencia*; es también miembro del Comité Científico del Archivio Scientifico Cassani de la Universidad Reggio Emilia y de la Colección "Derechos Humanos y Filosofía del Derecho" de la editorial Dykinson. Secretario de la colección "Panoramas de Derecho" de la Facultad de Derecho de la US con la editorial Thomson Reuters Aranzadi, Vice-Director de la revista *Crónica Jurídica Hispalense. Revista de la Facultad de Derecho de la Universidad de Sevilla* y Co-Director de la revista *Annaeus: Anales de la tradición romanística*; asimismo forma parte de los consejos de redacción de las revistas *Anuario de Filosofía del Derecho*, *The Age of Human Rights Journal*, *Derechos y Libertades*, *Persona y Derecho*, y *Cuadernos sobre Vico*. Ha publicado más de un centenar de trabajos de su especialidad, entre los que destacan sus monografías: *Homo Excelsior. Los límites ético-jurídicos del transhumanismo*, Tirant lo Blanch, Valencia, 2018; *Inteligencia Artificial y Derecho. El jurista ante los retos de la era digital* (coeditores: Fernando H. Llano Alonso y Joaquín Garrido Martín), Thomson Reuters Aranzadi, Cizur Menor (Navarra), 2021; recientemente ha publicado varios artículos que están directamente relacionados con la especialidad de este libro, como por ejemplo: "Transhumanism, vulnerability and human dignity", *Deusto Journal of Human Rights*, nº 4, 2019, pp. 39-58; y "L'etica dell'intelligenza artificiale nel quadro giuridico dell'Unione europea", *Ragion pratica*, nº 57, 2021, pp. 327-348.

LEONOR MORAL SORIANO

Profesora Titular de Derecho Administrativo en la Facultad de Derecho de la Universidad de Granada, donde es Vicedecana de Internacionalización. Es doctora en Derecho por el Instituto Universitario Europeo (Florencia); consiguió la beca Marie Skłodowska-Curie para investigar en la Universidad de Edimburgo junto al profesor Neil MacCormick; trabajó durante cuatro años

como investigadora Senior en el Instituto Max-Planck de Bienes Comunes de Bonn; y su última estancia profesional en el extranjero, durante cinco años, ha sido en la Representación de España ante la UE en Bruselas. Ha trabajado en equipos multidisciplinares junto a expertos en educación, politólogos y economistas. Sus intereses investigadores han transcurrido siempre entre varias disciplinas: Teoría del Derecho, Derecho Europeo, políticas educativas, y desde 2020, en Inteligencia Artificial. Ha publicado varios artículos en JCR Q1.

STEFANO PIETROPAOLI

Profesor Titular de Filosofía del Derecho en la Facultad de Derecho de la Universidad de Florencia (Italia), acreditado a Catedrático de Universidad. En 2006 se doctoró en Justicia Constitucional por el Departamento de Derecho Público de Pisa. Ha realizado actividades de investigación en las universidades de Florencia, Brescia y Camerino, y en el Instituto de Teoría y Técnicas de la Información Jurídica del CNR. En 2007, fue uno de los ganadores del Coloquio de Cortona de la Fundación Feltrinelli. Ha impartido clases y participado en conferencias en universidades, instituciones y centros de investigación italianos y extranjeros, como el Instituto Italiano de Estudios Filosóficos, la Universidad de Sevilla, la Universidad de Cantabria, la Fondazione Italiana del Notariato, la Universidad de Oviedo y la Université Nice Sophia Antipolis. Fue Vocal de la Comisión General de Doctorado de la Universidad de Murcia. Es miembro del Laboratorio Hans Kelsen de la Universidad de Salerno. Es miembro de la redacción y editor web de "Jura Gentium". Rivista di Filosofia del Diritto Internazionale e della Politica Globale", y miembro fundador del centro del mismo nombre. Miembro del consejo de redacción de la *Rivista Italiana di Informatica e Diritto*, del comité científico del Archivio storico-giuridico "Anselmo Cassani", del taller de informática sobre "Derecho, ética, tecnologías" del CRID - Centro di Ricerca Interdipartimentale su Discriminazioni e vulnerabilità (UNIMORE). Miembro de la Carl-Schmitt-Gesellschaft. Es miembro del Laboratorio Nacional de Ciberseguridad del CINI (Consorcio Interuniversitario Nacional de Informática). Recientemente coeditó junto a Thomas Casadei el libro: *Diritto e tecnologie informatiche. Questioni di informatica giuridica, prospettive istituzionali e sfide sociali*, Wolters-Kluwer-Cedam, Milano, 2021.

ÁLVARO SÁNCHEZ BRAVO

Profesor Contratado Doctor de Filosofía del Derecho en la Facultad de Derecho de la Universidad de Sevilla. Presidente de la Asociación Andaluza de Derecho, Medio Ambiente y Desarrollo Sostenible y responsable del Grupo de Investigación "Informática, Lógica y Derecho" de la Universidad de Sevilla. En su extensa obra científica cabe destacar sus trabajos sobre protección de datos,

Derecho medioambiental, Internet y sociedad de la información y el marco jurídico europeo de la Inteligencia Artificial. Entre sus principales publicaciones cabe destacar, entre otras, las siguientes monografías: *Internet y la sociedad europea de la información: implicaciones para los ciudadanos*, Universidad de Sevilla, 2001; *Derecho, Inteligencia Artificial y nuevos entornos digitales*, Punto Rojo, Sevilla, 2020; *Derechos humanos y transformación social* (directores: David Sánchez Rubio y Álvaro Sánchez Bravo), Dykinson, Madrid, 2021.

ADOLFO J. SÁNCHEZ HIDALGO

Profesor Contratado Doctor de Filosofía del Derecho en la Facultad de Derecho de la Universidad de Córdoba, acreditado a Profesor Titular de Universidad. Ha realizado estancias de investigación en las universidades: Paris II Pantheon Assas, Universidad de Catania y Universidad de Padua. Sus principales líneas de investigación son: a) Metodología jurídica, b) Historia de la Filosofía del Derecho, c) Derecho y Nuevas Tecnologías, d) Teoría Comunicacional del Derecho, y e) Derecho deportivo. Es miembro del grupo de investigación SEJ-050 Comunicación, Lenguaje y Derecho, forma parte de la Red Iberoamericana de Investigación en Gestión y Derecho del Deporte. Entre sus numerosas publicaciones destaca su monografía titulada: *Epistemología y metodología jurídica*, Tirant lo Blanch, Valencia, 2019.

Mª OLGA SÁNCHEZ MARTÍNEZ

Profesora Titular de Filosofía del Derecho en la Facultad de Derecho de la Universidad de Cantabria. Sus trabajos se centran en el ámbito de la Teoría del Derecho y Teoría de la Justicia, especialmente en relación a los Derechos Humanos, cuestiones de Género y Derecho y actualmente en temas referentes a las Nuevas Tecnologías, Derecho y derechos. Ha sido directora del Área de Igualdad de la UC y, desde octubre de 2019, es Directora del Departamento de Derecho Público. Por su trayectoria -docente, investigadora y de gestión- en materia de género ha sido merecedora de dos premios: Premio a la Igualdad en su IV edición, concedido por el Consejo de la Mujer del Gobierno de Cantabria, año 2011; y el Premio a la Igualdad de la Universidad de Cantabria, en su III edición, año 2014. En su amplia producción científica destaca, por su relación directa con la temática de este libro: "Desafíos democráticos en el ecosistema digital", *Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho* (editor: José Ignacio Solar Cayón), Universidad de Alcalá-Defensor del Pueblo, Madrid, 2020, pp. 79-119; "Las transformaciones de la libertad de expresión en las sociedades analógica y digital", *Manifestaciones contemporáneas del Derecho y los derechos humanos* (editora: Mª Isabel Garrido Gómez), Tirant lo Blanch, Valencia, 2021, pp. 267-298; "Tecnología digital e

Inteligencia Artificial: Nuevos retos y oportunidades para los derechos humanos", Teoría jurídica contemporánea, vol. 6, 2021, pp. 1-33.

MARÍA SEPÚLVEDA GÓMEZ

Profesora Titular de Derecho del Trabajo y de la Seguridad Social en la Facultad de Derecho de la Universidad de Sevilla. Entre sus publicaciones más recientes destacan, por su relación directa con la temática de esta obra colectiva: "Los derechos fundamentales inespecíficos a la intimidad y al secreto de las comunicaciones y el uso del correo electrónico en la relación laboral. Límites y contra límites", *Temas laborales. Revista andaluza de trabajo y bienestar social*, nº 122, 2013, pp. 197-214; "Poder de control y empresarial mediante cámaras de videovigilancia y derecho de los trabajadores a la protección de datos personales", Temas laborales. Revista Andaluza de Trabajo y Bienestar Social, nº 132, 2016, pp. 219-235; "Negociación colectiva y derechos digitales en el empleo público", *Revista General de Derecho del Trabajo y de la Seguridad Social*, nº. 54, 2019; "El acuerdo marco europeo sobre digitalización. El necesario protagonismo de la norma pactada", *Temas laborales: Revista andaluza de trabajo y bienestar social*, nº 158, 2021, pp. 213-244.

JOSÉ IGNACIO SOLAR CAYÓN

Profesor Titular de Filosofía del Derecho en la Facultad de Derecho de la Universidad de Cantabria, de la que ha sido Secretario General y, actualmente, Secretario del Consejo Social. Investigador principal del Proyecto de I+D+i "La inteligencia artificial jurídica" [RTI2018-096601-B-100 (MCIU/AEI/FEDER, UE)]. Está especializado en el estudio de la Inteligencia Artificial jurídica, la justicia algorítmica, el Derecho digital y la aplicación de las nuevas tecnologías al mercado jurídico. Entre sus publicaciones mas recientes destacan: "La codificación predictiva: inteligencia artificial en la averiguación procesal de los hechos relevantes", *Anuario de Filosofía del Derecho*, nº 11, 2018, pp. 75-105; *La Inteligencia Artificial jurídica. El impacto de la innovación tecnológica en la práctica del Derecho y el mercado de servicios jurídicos*, Thomson Reuters Aranzadi, Cizur Menor (Navarra), 2019; *Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho* (editor: José Ignacio Solar Cayón), Universidad de Alcalá-Defensor del Pueblo, Madrid, 2020; "Retos de la deontología de la abogacía en la era de la inteligencia artificial jurídica", *Derechos y libertades. Revista de Filosofía del Derecho y derechos humanos*, nº 45, 2021, pp. 123-161. "La inteligencia artificial jurídica: nuevas herramientas y perspectivas metodológicas para el jurista", *Revus. Journal for Constitutional Theory and Philosophy of Law*, nº 41, 2020, pp. 1-27. *El impacto de la inteligencia artificial en la teoría y la práctica jurídica* (editores: José Ignacio Solar Cayón y Mª Olga Sánchez Martínez), La Ley (Wolters Kluwer), Madrid, 2022.

RAMÓN VALDIVIA JIMÉNEZ

Doctor en Derecho por la Universidad de Sevilla y doctor en Filosofía por la Universidad Pontificia Lateranense. Profesor de Teoría y Filosofía del Derecho en el Centro de Estudios Universitarios "Cardenal Spínola", centro adscrito a la Universidad de Sevilla. Es sacerdote y Vicario episcopal de la Archidiócesis de Sevilla. Es miembro del Grupo de Investigación de la Junta de Andalucía SEJ-504 "Bioderecho Internacional". En 2021, su tesis doctoral, titulada: **"**El nacimiento de la modernidad: Justicia y Poder en Bartolomé de Las Casas (1484-1566)**"** obtuvo el Premio Internacional "Bartolomé de Las Casas", concedido por el Instituto para el Estudio de las Religiones y el Diálogo Interreligioso (IRD) de la Facultad de Teología de la Universidad de Friburgo. En su bibliografía, especializada en historia del pensamiento jurídico, libertad religiosa y ética de la Inteligencia Artificial, destacan, por su relación directa con la temática de este libro, los siguientes trabajos: "Revisión crítica del transhumanismo: Derecho a la vulnerabilidad en la esperanza cristiana"; *Ius et Scientia. Revista Electrónica de Derecho y Ciencia*, vol. 5, nº 1, 2019, pp. 265-281; "Ética e Inteligencia Artificial. Una discusión jurídica", *Ius et Scientia. Revista Electrónica de Derecho y Ciencia*, vol. 6, nº 2, 2020, pp. 111-134.

DIANA CAROLINA WISNER GLUSKO

Profesora Titular de Derecho Administrativo en el Centro de Estudios Universitarios "Cardenal Spínola", centro adscrito a la Universidad de Sevilla. Doctora en Derecho por la Universidad Carlos III de Madrid. Actualmente es Gestora del Área de conocimiento del Grado en Derecho. Coordinadora del Máster en Derecho Empresarial de las Nuevas Tecnologías del CEU Cardenal Spínola. Investigadora especializada en Administración electrónica y Derecho Administrativo y Empresarial en el ámbito de las Nuevas Tecnologías. Entre sus principales publicaciones destacan: "La confidencialidad de la información regulatoria en el sector de las telecomunicaciones", Revista General de Derecho Administrativo, nº 20, 2011, pp. 99-111; "Administración Electrónica inclusiva. Accesibilidad de los sitios web de los organismos del Sector Público", *Narrativas Sociopolíticas en pleno siglo XXI. Perspectivas multidisciplinares en un mundo global* (editores: D. L. Sutil Martín y A. Luna García), Global Knowledge Academics, 2017, pp. 113-130; "Accesibilidad de las sedes electrónicas. Hacia una administración electrónica inclusiva", *Administración electrónica. Retos jurídicos y tecnológicos de su implantación en Andalucía* (coordina: Diana Carolina Wisner Glusko), Fundación San Pablo Andalucía CEU, 2018, pp. 78-139.

Proyecto PID2019-108155RB-I00 financiado por MCIN/ AEI /10.13039/501100011033

ISBN 978-84-19145-25-3

Laborum
EDICIONES