

Álvaro Sánchez Bravo (Director)

Geraldo Costa Da Camino (Coordinador)

Intellegentiae Artificialis, Imperium et Civitatem



an

Álvaro Sánchez Bravo (Director)

Geraldo Costa da Camino (Coordinador)

INTELLENTIAE ARTIFICIALIS, IMPERIUM ET CIVITATEM

INTELLENTIAE ARTIFICIALIS, IMPERIUM ET CIVITATEM
VV. AA.

Editado por:
Editorial Alma Mater
Paseo de las Delicias, 30, 2º. Madrid (España)

info@editorialalmamater.com
www.editorialalmamater.com
Impreso en España
ISBN: 979-83-96042-25-4

Maquetación, diseño y producción: Editorial Alma Mater
© 2022 Editorial Alma Mater

Quedan rigurosamente prohibidas, sin la autorización por escrito de los titulares del copyright, bajo las sanciones establecidas por las leyes, la reproducción parcial o total de esta obra por cualquier medio o procedimiento, comprendidos la reprografía y el tratamiento informático, y la distribución de ejemplares de esta edición mediante alquiler o préstamos públicos.

ÍNDICE

LA TOMA DE DECISIONES AUTOMATIZADA Y EL CONTROL DE HUMANIDAD

Fernando H. Llano Alonso.....7

¿SE PUEDE AUTOMATIZAR LA *FAIRNESS* EN LA INTELIGENCIA ARTIFICIAL? LA PROBLEMÁTICA DE LA DISCRIMINACIÓN Y LA DIMENSIÓN DE LA INTERSECCIONALIDAD

Nuria Belloso Martín.....37

LA INTELIGENCIA ARTIFICIAL: EL NUEVO AVANCE TECNOLÓGICO CON REPERCUSIÓN EN LOS SISTEMAS JURÍDICOS

Susana Álvarez González..... 59

REGULAÇÃO DA INTELIGÊNCIA ARTIFICIAL NO BRASIL

Licurgo Mourão73

DE LAS ACTUACIONES ADMINISTRATIVAS AUTOMATIZADAS A LA INTELIGENCIA ARTIFICIAL. UN BREVE APUNTE SOBRE SU IMPLEMENTACIÓN Y REGULACIÓN EN LA ADMINISTRACIÓN PÚBLICA ESPAÑOLA

Elisa Rocío Prados Pérez..... 91

UM CONTROLE SOCIAL INOVADOR SOBRE O CONTROLE EXTERNO: MÉTODO PARA AFERIÇÃO DE DECISÕES DOS TRIBUNAIS DE CONTAS

Geraldo Costa da Camino. 111

CONTROLE DE SUSTENTABILIDADE (T)EC(N)OLÓGICO PELOS TRIBUNAIS DE CONTAS DO BRASIL E DA ESPANHA E UM BREVE DIÁLOGO COM CERVANTES

Daniela Zago Gonçalves da Cunda. 121

INTELIGENCIA ARTIFICIAL, ALGORITMOS Y TRABAJO

Juan Raso Delgue141

NOVAS PERSPECTIVAS PARA O CONTROLE
EXTERNO: ONLINE DISPUTE RESOLUTION,
IA E TRIBUNAIS DE CONTAS

Ricardo Schneider Rodrigues159

CIUDADES, TIC E INTELIGENCIA ARTIFICIAL

Álvaro Sánchez Bravo 175

INTELLENTIAE ARTIFICIALIS,
IMPERIUM ET CIVITATEM

LA TOMA DE DECISIONES AUTOMATIZADA Y EL CONTROL DE HUMANIDAD

Fernando H. Llano Alonso
Decano de la Facultad de Derecho
Universidad de Sevilla (España)

INTRODUCCIÓN

La relación entre el ser humano y la Inteligencia Artificial (IA) ante el horizonte de la singularidad ha alimentado tradicionalmente la imaginación de los autores que cultivan el género de la ciencia ficción y la distopía. En la mayoría de estas novelas distópicas son las máquinas inteligentes las que terminan controlando a la humanidad, ante la imposibilidad de ésta de competir con un adversario infinitamente superior en velocidad de procesamiento de datos, capacidad de almacenamiento de información, resistencia y precisión en la ejecución de tareas, y sobre todo en aprendizaje.

Cuando los humanos crearon un ordenador capaz de almacenar la información y aprender de ella, firmaron la sentencia de muerte de la humanidad.

Con esta cita se inicia *Dune. La yihad butleriana* (2004), primera novela de la trilogía de Brian Herbert y Kevin J. Anderson sobre *Leyendas de Dune*, que a su vez es el prelude de la saga *Dune* de Frank Herbert. Quizás sea esta la primera novela a contracorriente de la mayoría de novelas futuristas y distópicas en la que los últimos humanos libres se revelan en los Planetas Sincronizados, bajo el liderazgo de Serena Butler, contra el dominio de las máquinas pensantes desarrolladas con IA (todas ellas subsumidas en el supercomputador *Omnium*) y consiguen vencerlas rotundamente. Así describen Herbert y Anderson el comienzo del

fin del ciclo de dominio de las máquinas inteligentes y autoconscientes sobre los humanos:

Desde las profundidades de sus ciudades de energía, Omnium observaba la Tierra. Sus ojos espía grababan cada fase del audaz ataque humano. Vio que las tornas cambiaban (Herbert-Anderson, 2011⁵, 619)

A diferencia de otras novelas y películas de ciencia ficción, *Dune* supone una rara excepción a la resignación con la que los seres humanos aceptan la llegada de la era de la posthumanidad. Por ello, apelando precisamente al espíritu indómito de los últimos seres humanos autoconscientes y a su ansia de libertad, Serena Butler prende la llama de la rebelión contra las máquinas pensantes. Hay en esta novela una voluntad de resistencia colectiva de los seres humanos conscientes de su condición de sujetos morales y, por ende, dotados de conciencia moral, sentido de la responsabilidad y sabiduría práctica, de su derecho a la libertad y a la dignidad, a pesar de la opresión que sufren *de facto* a causa de la Singularidad tecnológica.

A propósito de la inmersión de la humanidad en el universo de la Singularidad se pregunta Slavoj Žižek si la entrada de la especie humana en el reino superior de lo posthumano supondrá la desaparición gradual de la humanidad tal y como la conocemos o si, por el contrario, nuestros futuros descendientes posthumanos que vivan en ese “estadio superior” de la perfección serán capaces de recordar de algún modo, incluso con una memoria entrañable, los azares del “estadio inferior” de la humanidad (Žižek, 2023, 142-143). En definitiva, ¿cómo se enfrentará la humanidad a la Singularidad (ese “gran otro” que, según Jacques Lacan, constituye el orden simbólico virtual integrado en la red digital), lo hará de manera consciente o inconsciente? En esa esfera de los múltiples metaversos ¿serán los humanos capaces de diferenciar la realidad virtual de lo real, las experiencias compartidas sincrónicamente

por todos los internautas neuroconectados a través de la IA de la experiencia subjetiva e individual vivida por cada persona?

Para ilustrar esta dicotomía en torno a la conservación del control de la conciencia subjetiva en el espacio infinito de la Singularidad donde se recrea una realidad que no existe pero que es simultáneamente compartida de forma colectiva, y donde no se sabe qué organismo regulador decidirá qué experiencias pueden compartirse ni quién controlará a su vez ese mecanismo, Žižek trae a colación la última escena de un film de ciencia ficción de culto: *Matrix* (1999). En dicha escena, Neo, el protagonista, un programador que actúa como pirata informático y que intuye que algo no va bien en el universo *virtual* en el que cree haber vivido *realmente* todo el tiempo, anuncia la liberación de la humanidad tras destruir Matrix, la realidad inducida por las máquinas a las mentes conectadas de hombres y mujeres alienados y sumidos en un letargo onírico. La paradoja de este final tan ambiguo de la película de las hermanas Wachowski es que, por un lado se anuncia a la humanidad su liberación del dominio de las máquinas en un mundo irreal, pero por otra parte se les arroja al “desierto de lo real”, un mundo destruido anteriormente por las máquinas (en el film se muestra concretamente la ciudad de Chicago reducida a escombros) en el que impera el vacío, la nada (Žižek, 2023, 84-85).

El recurso a la novela y el cine de ciencia ficción en la presentación de este último capítulo se debe en realidad a la necesidad de situar las conjeturas y elucubraciones del evolucionismo tecnológico y posthumanista fuera de los márgenes estrictamente científicos, es decir, el que se corresponde con los hechos probados. Recientemente Erik J. Larson se ha referido a la supuesta inevitabilidad de la Singularidad tecnológica en un libro que pretende desmontar el mito de la inexorabilidad e infalibilidad de la Inteligencia Artificial, que cuenta entre sus principales valedores a futurólogos transhumanistas como Ray Kurzweil y Nick Bostrom.

Según Larson, el peligro de tomarse en serio el futuro del mito de la IA es que puede deparar consecuencias muy negativas, ya que subvierte la ciencia. En este sentido, sostiene este científico experto en computación y emprendedor tecnológico

La ciencia de datos (la aplicación de la IA a los macrodatos) es, en el mejor de los casos, una prótesis del ingenio humano; en caso de usarla de manera correcta, nos ayudará a lidiar con el “diluvio de datos” contemporáneo. Cuando se la usa para reemplazar la inteligencia individual, tiende a estropear la inversión sin ofrecer ningún resultado (...) Estamos pagando un precio demasiado elevado por este mito. Como no poseemos ninguna buena razón científica para creer que el mito pueda hacerse realidad, puesto que contamos con todos los motivos para rechazarlo a fin de alcanzar la prosperidad en el futuro, tenemos que repensar de manera radical la conversación sobre la IA (Larson, 2022, 11).

Dejemos pues la ficción distópica aparte y vayamos a la realidad de la interacción humana con la automatización y las simulaciones interactivas con robots y sistemas de IA. Un claro ejemplo de la interacción remota entre el ser humano y la máquina de circuito cerrado lo proporciona el Internet táctil con el ser humano “en el bucle”, es decir, dentro del sistema de aprendizaje automático para ayudar a la computadora a tomar las decisiones correctas en la construcción de un modelo o en una simulación (*Tactile Internet with Human-in-the-Loop*).

En el siguiente epígrafe se presentarán sucintamente las cuatro variantes posibles que los expertos en robótica e ingeniería de telecomunicaciones distinguen a propósito de la interacción entre los humanos y las máquinas inteligentes, desde la que asegura el control humano de la máquina (*Human-in-the-Loop*), pasando por dos tipos intermedios de transferencia de habilidades y retroalimentación entre humanos y máquinas (*Human-in-the-Loop*).

for Exceptions y Human-on-the-Loop), hasta la cuarta modalidad, radicalmente opuesta a la primera, en la que se excluye al ser humano de la toma de decisiones automatizada de la máquina (*Human-out-of-the-Loop*).

CUATRO MODELOS DE INTERACCIÓN EN LA RELACIÓN ENTRE HUMANOS Y MÁQUINAS INTELIGENTES DENTRO DEL PROCESO DE TOMA DE DECISIONES

El proceso de digitalización y automatización vinculado al fenómeno de la revolución 4.0 ha metamorfoseado el modo tradicional de gestionar la toma de decisiones de las administraciones, de las empresas, de los servicios profesionales en todos los órdenes de la vida pública (sanitario, jurídico, educativo, científico-tecnológico...)

Imaginar un sistema autónomo y mecanizado de toma de decisiones resulta tan sencillo que no son pocos los psicólogos y especialistas en el estudio del comportamiento humano que coinciden en señalar cómo el uso de nuevos y sofisticados generadores de textos e imágenes impulsados por IA -por ejemplo, mediante herramientas como el Chat GPT- está afectando a la creatividad humana, el juicio y la toma de decisiones.

Ahora bien, desde el momento en que en el proceso de toma de decisión de una máquina inteligente interviene el ser humano, en la medida en que las consecuencias de la micro-decisión automatizada pueden afectarles (piénsese, por ejemplo, en un vehículo autónomo que ofrece a su conductor alternar las opciones de conducción manual o asistida), resulta imprescindible el diseño de un sistema de decisión que permita al ser humano tener una mínima interacción con la máquina (Ross-Taylor, 2021).

Los cuatro principales modelos de gestión que se han desarrollado para la interacción entre los humanos y las máquinas inteligentes varían en función del grado de implicación y de la naturaleza de la intervención humana: *Human-in-the-Loop* (HITL), *Human-in-the-Loop for Exceptions* (HITLFE), *Human-on-the-Loop* (HOTL), y *Human-out-of-the-Loop* (HOOTL).

A continuación se explicarán sucintamente cada una de estas variantes:

HUMAN-IN-THE-LOOP (HITL)

Este primer modelo de interacción con la máquina permite al usuario llevar el control en el proceso de aprendizaje automático (*machine learning*) y hacer una transferencia de competencias a la máquina o robot inteligente. En el ámbito de las simulaciones virtuales HITL, por citar algunos supuestos: simuladores de vuelo (por ejemplo, el X-62A Vista, utilizado por la Fuerza Aérea de los Estados Unidos), de conducción urbana (como el sistema español CARLA, acrónimo del inglés Car Learning to Act), o incluso de medicina “in silico” para probar medicinas y tratamientos experimentales en pacientes virtuales (uno de ellos, el simulador HeartFlow Analysis ayuda a localizar la enfermedad de las arterias coronarias a partir de las imágenes de tomografía computerizada del corazón de un paciente), los usuarios interactúan en tiempo real con la simulación a través de una interfaz gráfica, y se pueden recopilar directamente datos de los usuarios en un entorno experimental controlado (Rao, Chernyakhovsky, Rao, 2011, 157).

Se pueden elegir distintos caminos para intercambiar habilidades entre humanos y máquinas. Una vía para este tipo de transferencia de habilidades consiste en equipar a un experto humano con cualquier tipo de interfaz hombre-máquina que funciona como un simple mando a distancia (por ejemplo, el controlador de una videoconsola). Sin embargo, hay que tener en cuenta que no todos los expertos humanos son capaces de manejar un mando de este

tipo, ni entienden la relación entre los movimientos y la precisión del robot. Sin embargo, la industria de la robótica reconoce este problema de especialización humana desde hace tiempo, y por eso se establecen procesos estándar para transmitir conocimientos y experiencia a los robots industriales.

Cada máquina necesita ser controlada por un solo humano. Por eso, incluso en escenarios con millones de consumidores que demandan una habilidad específica del robot, el control humano debe ser escalonado a la hora de transferir las habilidades de un solo operador a millones de robots para satisfacer adecuadamente la demanda de los consumidores (Fitzek-Li-Speidel-Strufe, 2021, 4-5).

Pero la robótica actual también permite un *feedback* o retroalimentación multimodal que permite la transferencia a los humanos de competencias aprendidas automáticamente por la máquina a través del proceso de *machine learning*. En esta transferencia de habilidades en sentido inverso es la máquina quien enseña al humano; pensemos, por ejemplo, en los *wearables* o dispositivos inteligentes que portamos a diario, como pulseras, gafas o relojes, y que al estar interconectados a Internet permiten conocer con precisión datos como nuestra localización exacta en tiempo real, nuestro ritmo cardiaco, los kilómetros que hemos recorrido o las calorías consumidas.

Suponiendo que los *wearables* no sólo estén equipados con sensores sino también con *actuadores* (dispositivos mecánicos esenciales que sirven para mover o actuar sobre otros dispositivos mecánicos), las señales de aprendizaje pueden generarse en directo o con antelación y transmitirlas al usuario humano. En la rehabilitación física, por ejemplo, los movimientos de un fisioterapeuta a distancia podrían generarse en línea (es decir, a distancia) y transmitirse a los *wearables* equipados con actuadores que lleve una persona mayor con movilidad reducida, o a un dispositivo ciberfísico que ayude a realizar ejercicios de

fisioterapia a los pacientes en casa. Los posibles ámbitos de aplicación de esta transferencia de habilidades no se limitan tan solo a los cuidados sanitarios y de enfermería, sino que también comprenden la enseñanza de nuevas habilidades en la escuela, el trabajo o los intereses personales, es decir, que se encuentran dentro del amplio espectro del Internet de las Habilidades.

HUMAN-IN-THE-LOOP-FOR-EXCEPTIONS (HITLFE)

La mayoría de las decisiones están automatizadas en este segundo modelo, y el humano sólo se ocupa de las excepciones. Para las excepciones, el sistema requiere algún juicio o aportación del humano antes de tomar la decisión. Además, los humanos también controlan la lógica del sistema para determinar cuáles serían las excepciones susceptibles de revisarse en el caso de que su intervención fuera necesaria, por ejemplo, cuando se pone en cuestión el nivel de confianza en las predicciones de un sistema predictivo de IA basado en algoritmos de aprendizaje automático.

En relación con la falibilidad de los sistemas de algoritmos predictivos, un caso paradigmático es el de Glossier, una marca de belleza online que desarrolló un algoritmo de aprendizaje automático para predecir el comportamiento de sus clientes y su compromiso futuro en el aumento de ventas de distintas promociones de sus productos. Sin embargo, la predicción *machine learning* asumió erróneamente que las visitas a su sitio web “Into the Gloss” de los potenciales clientes se traducirían necesariamente en compras online teniendo en cuenta factores como la oferta, el apoyo de marketing y la estacionalidad para crear una previsión automatizada. Y en efecto, para muchas promociones la predicción *machine learning* funcionó bien, pero los directivos perdieron rápidamente la confianza después de que, tras su éxito inicial, se produjera una serie de estrepitosos fracasos predictivos que se tradujeron en importantes pérdidas de ventas.

Cuando los científicos de datos revisaron las predicciones algorítmicas basadas en los datos extraídos de los millones de datos y patrones de comportamiento de sus clientes en las plataformas digitales y redes sociales, descubrieron que el algoritmo de aprendizaje automático tenía dificultades para predecir ciertos tipos de promociones. Pero, en lugar de abandonar el proyecto, los científicos decidieron desarrollar un enfoque HITLFE. La clave consistía en codificar el nivel de confianza de la máquina en sus predicciones y hacer que los humanos revisaran dichas predicciones de forma excepcional cuando la máquina no les ofreciera la fiabilidad esperada (Ross-Taylor, 2021).

HUMAN-ON-THE-LOOP (HOTL)

En el tercer modelo de interacción entre humanos y máquinas inteligentes o robots, son los humanos quienes asisten a las máquinas en su toma de microdecisiones mecánicas. En los sistemas de HOLP la toma de decisiones debe automatizarse al 100% porque la respuesta debe producirse casi de manera inmediata, en un periodo de tiempo demasiado corto como para incluir en el sistema al ser humano. Sin embargo, aunque el operador humano quede excluido en primera instancia de ese proceso de toma de decisión automatizada, sí puede supervisar posteriormente los resultados de estas decisiones automatizadas e incluso ajustar reglas y parámetros para futuras decisiones. En una configuración algorítmica más avanzada, la máquina también recomienda parámetros o cambios en las reglas que luego son supervisados por un humano.

Un ejemplo de aplicación del sistema autónomo de toma de decisiones HOTL lo encontramos en una aseguradora global de vida y salud que necesita aumentar el valor de sus ventas. En este supuesto, los agentes de ventas de esta empresa utilizan un dispositivo móvil para trabajar con posibles clientes, recopilando datos sobre sus necesidades financieras y su situación, les y ayudan

a solicitar electrónicamente productos de protección e inversión. El aumento de los ingresos resultará de gran utilidad a los agentes para identificar las oportunidades adecuadas de venta cruzada y venta adicional mientras tratan con los clientes potenciales.

Posteriormente, el departamento de marketing de esta compañía aseguradora utilizará análisis avanzados e IA para identificar ofertas para cada cliente potencial en función de su situación financiera y de los productos que compre. Este sistema de toma de decisiones a nivel de cliente se integra perfectamente en la aplicación móvil de la aseguradora y se ejecuta en tiempo real, de forma totalmente autónoma. Luego el equipo de marketing gestionará el bucle de retroalimentación y obtendrá datos detallados sobre cómo se seleccionan las ofertas, podrá hacer un seguimiento de las estrategias de los clientes y realizar cambios en el comportamiento del sistema.

Por cierto, según algunos estudios recientes sobre el impacto de la IA en el futuro del sector de los seguros en la próxima década (2030), las aseguradoras que prosperarán más serán precisamente aquellas que mejor utilicen las nuevas tecnologías para crear productos innovadores, aprovechar los conocimientos del aprendizaje cognitivo a partir de nuevas fuentes de datos, agilizar los procesos y reducir los costes, y superar las expectativas de los clientes en cuanto a individualización y adaptación dinámica (Balasubramanian-Libarikian-McElhaney, 2021).

HUMAN-OUT-OF-THE-LOOP (HOOTL)

En el cuarto modelo de interacción, radicalmente opuesto al primero (HITL), la máquina es quien toma todas las microdecisiones, aunque ésta es supervisada desde fuera del sistema por el hombre que interviene tan solo para establecer nuevos límites y objetivos. En esta modalidad el *feedback* del sistema se encuentra dentro de un bucle cerrado, por lo que los ajustes de la

retroalimentación de datos entre los humanos y las máquinas se hallan automatizados.

A propósito del sistema de toma automatizada de decisiones HOOTL, el *Mayflower Autonomous Ship*, un buque completamente autónomo (no tripulado) de investigación marina cuyo software ha sido co-diseñado por la compañía IBM y la organización sin fines de lucro Promoting Marine Research and Exploration (ProMare), recopila a lo largo de su travesía cantidades masivas de datos, incluso en entornos meteorológicos hostiles, y tiene capacidad de tomar decisiones cruciales en fracciones de segundo sin intervención humana de ningún tipo. La capitana de la nave es la IA que navega autónomamente, surca los mares evitando los peligros oceánicos y observa rigurosamente las reglas del Derecho Marítimo, con el fin de alcanzar los objetivos preestablecidos por los responsables del proyecto científico (que son seres humanos).

Respecto a los cuatro modelos de interacción que se han explicado de forma sucinta solo cabría añadir que, con independencia de cuál sea la opción de gestión con la que deseemos interactuar con la IA (ya sea desde la mayor implicación humana en la supervisión algorítmica del HITL, a la completa autonomía de la máquina en la toma de decisiones propia del HOTL), todo sistema de micro-decisión debe ser revisado para garantizar que la toma de decisión automatizada es la adecuada, para ello es fundamental que se eviten las decisiones herméticas que se amparan en la opacidad de los algoritmos de cajas negras, y que ninguna decisión algorítmica que se aplique al mundo real pueda optimizarse en tanto que esté basada en una sola métrica, sin que exista una mínima supervisión y una compensación entre los parámetros que sirven para medir la toma de decisiones.

En el próximo epígrafe me referiré precisamente a la transparencia algorítmica y a la necesaria garantía de la reserva y el control de humanidad en materia de libertades y derechos fundamentales.

LA RESERVA DE HUMANIDAD Y EL CONTROL DE LOS ALGORITMOS EN MATERIA DE LIBERTADES Y DERECHOS FUNDAMENTALES

En la doctrina iuspublicista española más reciente ha alcanzado especial fortuna la expresión “reserva de humanidad”, una idea comparable a otros conceptos normativos afines como el de “reserva de ley” (por ejemplo, la reserva de ley orgánica que establece el art. 81.1 CE para el desarrollo de los derechos fundamentales y de las libertades públicas), o el término “reserva de ejercicio de potestades para los funcionarios”, según se prevé en el art. 9.2 del Real Decreto Legislativo 5/2015, de 30 de octubre, que considera que las potestades administrativas sólo deben ser ejercidas por los empleados públicos, debido a su relación estatutaria con la Administración, actúan al servicio los intereses generales de acuerdo con los principios de imparcialidad y objetividad.

En el ámbito reglamentario de la UE también se regula la reserva de humanidad por el Reglamento 2016/679 del Parlamento Europeo y del Consejo 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 (UE RGPD), en cuyo art. 22.1 se garantiza a quienes estén interesados el derecho a no ser objetos de decisiones individuales automatizadas, incluida la elaboración de perfiles, que produzca en ellos efectos jurídicos o les afecten significativamente. Por otra parte, en el art. 22.3 UE RGPD se introduce el control de humanidad (la supervisión humana) sobre la decisión automática para salvaguardar los derechos, las libertades y los intereses legítimos de los interesados.

Sin embargo, como ha señalado Juli Ponce, la inminente entrada en vigor del Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de Inteligencia Artificial (en adelante RUEIA) plantea problemas de coherencia normativa en la interpretación del art. 22 UE RGPD,

que establece una reserva de humanidad *generalizada*, salvo las excepciones previstas por la norma jurídica y su sustitución por la supervisión humana, y los arts. 5 y 14 RUEIA, que establecen, respectivamente, la prohibición de la IA *solo* en ciertos casos y la supervisión humana *únicamente* en supuestos de sistemas de IA de alto riesgo; por lo demás, en el resto de usos de la IA que no implican alto riesgo, el RUEIA no prevé ni reserva de humanidad ni supervisión humana (Ponce Solé, 2022, 66).

En el décimo punto de la Exposición de Motivos del RUEIA se indica que los Estados miembros deben estar facultados para especificar la aplicación de las disposiciones del RUEIA a través de su propia normativa nacional, para lo cual el Reglamento concede a los Estados un margen de maniobra en la concreción de dicha normativa interna, incluso en el tratamiento de datos personales (“datos sensibles”).

También en el párrafo 16 de la Resolución del Parlamento Europeo, de 3 de mayo de 2022, sobre Inteligencia Artificial en la era digital se encomienda a los legisladores el deber de “abordar los riesgos que plantea actualmente la toma de decisiones basada en la IA”, porque, como corresponde a un Estado de Derecho y a los principios proclamados en el art. 9.3 CE que le son consustanciales, el uso de la IA por las administraciones públicas no escapa al cumplimiento de los principios de legalidad, publicidad de las normas, seguridad jurídica, responsabilidad e interdicción de la arbitrariedad de los poderes públicos (Presno Linera, 2022, 51).

En este sentido, en relación con la legislación española sobre IA y la toma de decisiones administrativas, en el marco de la Estrategia Nacional de la Inteligencia Artificial, de la Carta de Derechos Digitales, y de las iniciativas europeas en torno a la IA, el art. 23 de la Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación, aparte de ser considerada por algunos juristas como la primera regulación de la IA en Derecho español, establece los principios de buena administración y diligencia debida como

límites de las decisiones administrativas automatizadas (Ponce Solé, 2019a; 2019b).

De conformidad con este estándar jurídico, las administraciones públicas tendrán en cuenta criterios de minimización de sesgos, transparencia y rendición de cuentas, y, siguiendo las recomendaciones de la UE, promoverán “el uso de una IA ética, confiable y respetuosa con los derechos fundamentales”. A hilo de este razonamiento sobre la pertinencia de la justificación o explicación de las decisiones públicas, y considerando los riesgos derivados del funcionamiento técnico de los algoritmos y de la IA empleados por la administraciones públicas, comenta Andrés Boix Palop que un mínimo de explicabilidad de los algoritmos es exigible, aún en el caso de que entendamos funcionalmente a los algoritmos como actuaciones administrativas y no como normas reglamentarias, de conformidad con lo establecido en los arts. 9.3 CE y 35 Ley 39/2015, de Procedimiento Administrativo Común (Boix Palop, 2022, 99).

La mención a la necesidad de avanzar en la rendición de cuentas de los programas informáticos con capacidad para tomar decisiones algorítmicas resulta oportuna, no solo en aras de la protección de los derechos y libertades de los ciudadanos, tanto en el ámbito de lo público como en el de lo privado, sino también porque, como ha advertido Manuel Medina Guerrero, a pesar de las innegables ventajas del empleo de sistemas de decisiones automatizadas basadas en algoritmos en términos de eficacia y economía, no pueden soslayarse los riesgos que conllevan estas decisiones para los derechos e intereses de la ciudadanía,

riesgos que se condensan en el temor genérico a que el ser humano se convierta en un mero objeto de los programas informáticos (Medina Guerrero, 2022, 141).

Algunos autores plantean ya la conveniencia de amparar por vía constitucional el derecho de los ciudadanos al conocimiento de los

algoritmos utilizados por las administraciones en la toma de decisiones automatizadas en el sector público. La justificación de la ampliación del reconocimiento constitucional a este pretendido derecho se motivaría por vía del art. 18.4 CE, en el que se consagran como derechos fundamentales la libertad informática y la protección de datos personales. Por cierto, esta interpretación abierta del art. 18.4 para integrar el derecho a conocer los algoritmos usados en la toma de decisiones parece haber encontrado receptividad en el legislador a través del derecho de acceso a la información pública regulado en la Ley 19/2013, de 9 de diciembre, y en la jurisprudencia constitucional reciente (STC 292/2020) a propósito de las facultades y poderes jurídicos que confiere al titular del derecho a la protección de datos (Cerrillo i Martínez, 2021, 41-78; Medina Guerrero, 2022, 142).

INTELIGENCIA ARTIFICIAL Y EL COTO VEDADO DE LA HUMANIDAD

En la toma de decisiones automatizadas hay una dimensión fundamental que debe tenerse en cuenta en la valoración del modelo de interacción entre humanos y máquinas, y es la perspectiva ético-jurídica desde la que se estudia el impacto de la cibernética, la informática, la IA y la robótica en el ámbito de los principios y los valores, un terreno de común interés para la Ética de la Inteligencia Artificial y la Filosofía del Derecho.

El pionero de los estudios de informática jurídica e iuscibernética ha sido, precisamente, un iusfilósofo: Mario G. Losano, quien en uno de sus primeros libros, *Giuscibernetica* (1969), analizaba la estructura interdisciplinar de la iuscibernética para ofrecer una síntesis adecuada de los principales problemas teóricos y prácticos que se encontraban los programadores y los juristas en el uso de los programas informáticos (problemas de índole sociológico, filosófico, lógico, lingüístico y técnico).

Por cuanto respecta a la idea de reserva de humanidad fue apuntada por primera vez, en sentido iusfilosófico, por Ernesto Garzón Valdés. En efecto, en uno de los primeros números de la revista *Doxa. Cuadernos de Filosofía del derecho* el iusfilósofo argentino publicó dos artículos; en el primero de ellos, titulado “Representación y democracia”, el iusfilósofo argentino se refiere por primera vez al “coto vedado” de los bienes básicos, que son aquellos que son condición necesaria para la realización de cualquier plan de vida (Garzón Valdés, 1989b, 209), y justifica el ejercicio del paternalismo jurídico por parte del Estado de derecho en el caso de que los miembros de la comunidad no comprendan la importancia de estos bienes básicos.

Por lo tanto, según Garzón Valdés, el coto vedado de los bienes básicos no dependería, en última instancia, de la voluntad o deseos de la comunidad, porque, en su opinión:

Quien no comprende la relevancia de los bienes básicos puede ser incluido en la categoría de incompetente básico (Garzón Valdés, 1989a, 157).

Para este autor, el rechazo de la garantía de los propios bienes básicos supone una muestra evidente de irracionalidad o de ignorancia de relaciones causales elementales como las que existen entre la disponibilidad de estos bienes y la realización de cualquier plan de vida.

Por otra parte, señala Garzón Valdés, los derechos y necesidades básicas incluidos en el coto vedado deben ser universalizables e iguales para todos los ciudadanos, pues sólo se puede concluir que una sociedad es homogénea cuando todos sus miembros gozan de los derechos incluidos en el coto vedado de los bienes básicos. Desde un punto de vista jurídico-positivo, el coto vedado de los intereses universalizables o derechos humanos no pueden ser objeto de recortes resultantes de negociaciones parlamentarias, en la medida que éstos constituyen

el núcleo no negociable de una constitución democrático-liberal que propician el Estado social (Garzón Valdés, 1989a, 162).

Existe una vinculación conceptual entre el coto vedado de los derechos, las necesidades y bienes básicos, y la democracia representativa, en la que está vigente el principio de la mayoría, dentro de un marco de homogeneidad social compatible con la realización de la esperanza de la autodeterminación individual. En definitiva, como sostiene Carlos Santiago Nino, en sintonía con Ernesto Garzón Valdés, el concepto de bienes y necesidades básicas

no sólo sería central en una concepción liberal de la sociedad, sino que haría de puente -al permitir su satisfacción simultánea- entre las dos ideas básicas del liberalismo: la de que los fines de los individuos deben ser respetados y la de que todo individuo es un fin en sí mismo (Nino, 1989, 34).

Garzón Valdés se refiere al coto vedado de los derechos, bienes y necesidades en clave liberal y neokantiana. En este sentido, conviene aclarar que su aproximación al concepto de bien básico desde un punto de vista moral recuerda a la idea del bien primario en la posición original rawlsiana e implica la ejecución exitosa de un plan de vida racional. Esta concepción de los bienes básicos (*primary goods*) permite a los individuos presumir la consecución de sus fines: libertad y oportunidad, renta y patrimonio, y sobre todo, el respeto de la persona por sí misma (*self-respect*) (Rawls, 1971, 433).

En línea de continuidad con esta lectura kantiana y liberal de los derechos fundamentales (en forma de lista de capacidades) Martha C. Nussbaum también es partidaria de un “blindaje” de los mismos, por ejemplo, ante cualquier intento de recorte o frente a posibles cambios que pudieran ser decididos por mayoría simple

en un nuevo proceso constituyente (Nussbaum, 2017, 69; 2012, 97-98).

Proyectando la tesis iusfilosófica del coto vedado defendida por Garzón Valdés al contorno ético-jurídico demarcado por la reserva de humanidad, en la que se protegen derechos, bienes y necesidades básicas de los ciudadanos frente la indeseada intromisión de las decisiones automatizadas basadas en algoritmos de alto riesgo, cabría preguntarse qué áreas prioritarias de los derechos y libertades serían aquellas en las que operaría este ámbito de protección de lo que es esencialmente humano, absolutamente irreductible e irrenunciable para la dignidad de la persona y la conciencia humana, ante la irrupción de la IA, la robótica y las tecnologías conexas.

En el art. 14 de 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 Unión¹, se regula la vigilancia humana de los sistemas de IA de alto riesgo, que -según la definición del texto normativo- son aquellos que comprometen la salud, la seguridad o los derechos fundamentales, y que pueden surgir

cuando un sistema de IA de alto riesgo se utiliza conforme a su finalidad prevista o cuando se le da un uso indebido razonablemente previsible (art. 14.2).

En los últimos años ha habido casos paradigmáticos y problemáticos derivados del uso de algoritmos de alto riesgo que han afectado precisamente a las tres grandes áreas a las que se refiere el mencionado precepto de la propuesta de Reglamento UE

¹ Texto de la Comisión Europea COM (2021) 206 final, publicado el 21 de abril de 2021, 2021/0106 (COD).

de IA. Veamos un ejemplo representativo por cada una de estas áreas:

En relación con la seguridad: el miércoles 19 de diciembre de 2018, a las 21:03 hrs, el aeropuerto de Gatwick se vio obligado a cerrar sus vuelos de entrada y salida durante 32 hrs debido al avistamiento cerca del aeródromo de varios drones. Más de 120.000 personas se vieron afectadas por este cierre en el segundo aeropuerto londinense (Shackle, 2020). Tan solo unas semanas más tarde, el 8 de enero de 2019, el aeropuerto de Heathrow, el primero en volumen de tráfico aéreo del Reino Unido, también suspendió durante una hora sus vuelos tras detectar la torre de control, a las 17.12 hrs, un dron sobrevolando su espacio aéreo. Desde un punto de vista legal y judicial estos incidentes generaron cambios (el 31 de diciembre de 2020 entró en vigor una nueva normativa para propietarios de drones, clasificando tipos de drones en diversas categorías, introduciendo requisitos de registro y seguro para cada clase). Por otra parte, se suscitó la cuestión a propósito de las reclamaciones de indemnización a las compañías de seguros por parte de los viajeros, y determinar sobre quién recaería la responsabilidad civil en ambos incidentes, se presentaría también la ocasión de plantearse la oportunidad de reconocer la personalidad electrónica de los robots o máquinas inteligentes entrenadas con IA para responder por los perjuicios causados a terceros por su mal funcionamiento o por sus erróneas tomas de decisión automatizadas (Benon-Decker, 2021).

Respecto a la reserva de humanidad y al control de las decisiones administrativas basadas en algoritmos predictivos cuando limitan el disfrute de los derechos fundamentales, en el ámbito de la educación (art. 27 CE) resulta ilustrativo el caso de los exámenes de acceso a las universidades británicas en el verano de 2020, el primero de la pandemia del Covid-19, cuyos resultados no se obtuvieron a partir de una prueba presencial, sino excepcionalmente mediante un algoritmo que combinó la calificación obtenida por los alumnos en sus colegios o institutos en los exámenes realizados en los tres últimos cursos y una

calificación estimada de cada alumno en comparación con sus compañeros de su centro.

El resultado de las predicciones del algoritmo utilizado para determinar las notas GCSE y A-Levels de los alumnos de Inglaterra, Gales y Escocia fue, anunciado el 13 de agosto, fue decepcionante: las calificaciones asignadas por el algoritmo fueron, en general, inferiores a las previsiones realizadas por los profesores de los colegios e institutos; por otra parte, se apreció también en la decisión automatizada un sesgo discriminatorio hacia los alumnos de las escuelas públicas situadas en zonas humildes y que, por el contrario, favorecía a los estudiantes de colegios privados o centros públicos de alto rendimiento. A raíz de la polémica causada por la prueba de acceso al sistema universitario británico mediante el empleo de un algoritmo predictivo, el gobierno anunció el 17 de agosto que los resultados del nivel A se modificarían para introducir en ellos como criterio corrector las estimaciones originalmente realizadas por los profesores de los centros de procedencia de cada alumno (por cierto, como ha podido apreciarse, este caso encajaría en el segundo modelo de control de humanidad: HITLFE).

A propósito de los algoritmos predictivos, se pregunta Alejandro Huergo Lora si en nuestra experiencia jurídica se pueden utilizar para determinar el contenido de resoluciones administrativas, al igual que sucediera con los A-Levels en 2020. En este sentido, el catedrático de Derecho Administrativo de la Universidad de Oviedo establece la siguiente diferenciación:

En el caso de potestades discrecionales, en las que la norma no vincula el contenido de la decisión administrativa a la constatación de uno o varios hechos, sino que establece un marco dentro del que es válida cualquier decisión que tome la Administración siempre que esté adecuadamente motivada y se hayan respetado las normas procedimentales, es perfectamente posible y válido que la

Administración utilice, entre otros factores, predicciones algorítmicas (Huergo Lora, 2020, 78-79).

Ahora bien, matiza Huergo Lora, no se puede soslayar que el hecho de que una decisión administrativa se constante mediante un modelo algorítmico predictivo plantea problemas de principio (en la medida en que, a su parecer, resultaría incoherente que se pueda decidir en función de predicciones y no de hechos probados) y problemas de orden práctico (¿cómo se pueden tomar decisiones a partir de algoritmos dada su opacidad?).

Teniendo en cuenta que el riesgo de error en la predicción algorítmica es aún muy alto, y que, dadas las circunstancias de error y opacidad de los algoritmos, el empleo de dichas predicciones en las resoluciones administrativas sería incompatible con las ideas de dignidad de la persona y del libre desarrollo de la personalidad, porque dicha predicción se basaría en la presunción de que los comportamientos que en el pasado han tenido otras personas en circunstancias similares a las del interesado que invoca ahora la decisión automatizada de la Administración por vía de predicción algorítmica, permitirían predecir con seguridad la conducta futura de ese interesado (Huergo Lora, 2020, 79).

Por último, respecto a la problemática que se origina a partir de la aplicación de la los sistemas expertos de IA y los algoritmos en el ámbito de la salud, conjugando su empleo con el marco legal que protege los derechos de los pacientes (consentimiento informado, privacidad y confidencialidad de los datos médicos, acceso al historial clínico, respeto de la voluntad anticipada por el paciente respecto a los cuidados y tratamientos de salud que desea recibir, derecho a presentar reclamaciones y sugerencias²).

² En España existe un amplio marco legal regulador de los derechos y obligaciones de los pacientes conformado por la siguiente normativa: el art. 43 CE, donde se reconoce el derecho a la protección de la salud; la Ley 14/1986, de 25 de abril, en cuyo art. 10. en el que se recogen los derechos de los pacientes con respecto a

En los últimos años se ha producido un aumento significativo del interés por las aplicaciones de los algoritmos de aprendizaje automático (*machine learning algorithms*) para la toma de decisiones médicas automatizadas. En los sistemas expertos aplicados a la diagnosis médica funcionan de la misma forma que el resto de sistemas expertos de la IA clásica, en la medida que contienen una base de datos de reglas deductivas mediante las cuales, a partir de un conjunto de hechos conocidos, se pueden inferir determinadas consecuencias. Concretamente, en medicina, la presencia de determinados síntomas puede hacer que el sistema experto los asocie a determinadas enfermedades relacionadas con dichos síntomas y sugiera realizar pruebas diagnósticas para llegar a conclusiones inequívocas.

Los sistemas de IA empleados en las diagnosis médicas están basados en conocimientos codificados de los expertos, como el sistema IBM Watson Oncology para la diagnosis y el tratamiento oncológico, un sistema de IA de metaanálisis que combina la extracción automática de textos de documentos clínicos con un número ingente de reglas lógicas. En el diagnóstico médico basado en imágenes, un sistema experto como éste es capaz de buscar características definidas por los expertos y codificar explícitamente las reglas de decisión definidas por los médicos (Jie-Zhiying-Li, 2021).

En suma, los algoritmos de aprendizaje automático han transformado para bien la atención sanitaria profesional, sobre todo si se compara el porcentaje de acierto y precisión de los diagnósticos realizados por los sistemas expertos de IA con los diagnósticos realizados por médicos especializados. En este sentido, según un estudio publicado en 2015 por la *National*

las distintas administraciones públicas sanitarias; la Ley Básica Reguladora de la Autonomía del Paciente y de Derechos y Obligaciones en Materia de Información y Documentación Clínica (Ley 41/2002, de 14 de noviembre); la Ley 16/2003, de 28 de mayo, de Cohesión y Calidad del Sistema Nacional de Salud, y el Real Decreto-Ley 7/2018, de 27 de julio, sobre el acceso universal al Sistema Nacional de Salud.

Academy of Sciences, alrededor de un 5% de los diagnósticos médicos realizados a adultos en centros hospitalarios estadounidenses fueron erróneos; a su vez, se estimaba que un 10% de los fallecimientos de los pacientes se debieron precisamente a errores de diagnóstico médico³.

En descargo de los responsables de los diagnósticos médicos realizados por clínicos humanos hay que remarcar el hecho de que, a diferencia de los sistemas expertos de IA, los clínicos, en cuanto humanos, son conscientes de su falibilidad y de la responsabilidad que asumen ante sus pacientes; por el contrario, existe una presunción favorable hacia la diagnosis mecánica realizada a través del algoritmo de aprendizaje automático, en la medida que ofrece mayor certeza y precisión (aunque en realidad haya también un margen de error en los algoritmos de aprendizaje automático). Pero la diagnosis médica plantea un problema aún mayor que el del cálculo meramente estadístico y cuantificativo de la exactitud de los resultados de las diagnosis clínicas: se trata de un problema epistémico-ético.

En un artículo reciente sobre la ética de la toma de decisiones algorítmica en la atención sanitaria sus autores, Thomas Grote y Philipp Berens, reconocen la utilidad de los algoritmos de aprendizaje automático para mejorar la capacidad de decisión de los clínicos al aportarles una fuente adicional de información y pruebas complementarias que les ayudarán a tomar sin duda una mejor decisión médica.

Sin embargo, los clínicos se enfrentan a un serio obstáculo cada vez que intentan inferir información a partir de los resultados de un algoritmo de aprendizaje automático. El problema subyacente puede describirse del siguiente modo: tanto el médico como el algoritmo de aprendizaje automático pueden considerarse expertos. Sin embargo, han recibido una formación diferente y

³ Cfr., *National Academies of Sciences*, "Engineering, and Medicine. Improving Diagnosis in Health Care", The National Academies Press, Washington DC, 2015.

razonan de formas muy distintas. Para el clínico, esta diversidad de formación y razonamiento plantea serios inconvenientes deontológicos y epistemológicos siempre que se produce un desacuerdo entre colegas o pares médicos considerados iguales (incluso cuando se reconoce como competentes a los algoritmos en igualdad de condiciones que sus “pares médicos”).

Por lo tanto, en primera instancia, en el caso de "desacuerdo entre iguales" de dos colegas o pares se produciría una discrepancia en la diagnosis de los síntomas, la historia clínica y el examen físico del paciente. Esta discrepancia entre “pares” (clínico y algorítmico) podría inclinarse a favor del algoritmo de aprendizaje automático por simple desistimiento del médico para eludir ulteriores responsabilidades derivadas de una toma de decisión errónea; con lo cual, desde un punto de vista epistémico y ético, además de fomentarse entre los médicos el dogmatismo y la credulidad científicista-tecnologicista, terminaría imponiéndose, antes que el mejor criterio médico, un simple mecanismo pragmático de “medicina defensiva” que protegería a los clínicos de la posible rendición de cuentas a sus pacientes (Grote-Berens, 2020, 208).

A veces la presunción de perfectibilidad e inexorabilidad del diagnóstico determinado por el algoritmo de aprendizaje autónomo puede ser la causa del desistimiento del control de humanidad por parte de los médicos respecto al funcionamiento correcto del sistema de IA aplicado al ámbito de la salud. En este sentido conviene no rebajar la exigencia del enfoque ético de la IA, pues de producirse esa renuncia por parte de la comunidad médica a la supervisión de las tomas de decisión automatizadas, se estaría abandonando (y sacrificando) injustificadamente el primer modelo de gestión para la interacción entre los humanos y las máquinas inteligentes (HITL) en el que el ser humano conserva el control sobre el sistema algorítmico, en beneficio del modelo antagónico (HOTL), en el que la decisión médica se confiaría, en última instancia, a las máquinas mientras que los clínicos se inhiben de participar en la misma para no asumir responsabilidades. Esta renuncia supondría, en última instancia, una rendición por parte los

profesionales de la medicina al mito de la perfección de la IA, además de su efectiva rendición a la falacia tecnológica inspirada por el fundamentalismo tecnológico, que no puede confundirse con la metodología empírica, la capacidad crítica y la vocación humanista que caracteriza a la ciencia médica.

En el próximo apartado, el conclusivo, se abordará precisamente esta cuestión relativa a la necesidad de supervisión ética de los algoritmos y la exigencia de responsabilidad derivada del funcionamiento erróneo o del mal uso de que pueda hacerse de los mismos.

CONCLUSIÓN

En 2016 Google presentó el proyecto DeepMind Health, en colaboración con el Royal Free London NHS Foundation Trust. En este proyecto de la multinacional tecnológica estadounidense puso en funcionamiento dos aplicaciones: Streams y Hark, con las que pretendía mejorar los sistemas de salud (en particular el NHS, es decir, el servicio de salud británico). Así, mientras que la app Streams servía para detectar pacientes en riesgo de contraer alguna enfermedad y se reducía a unos segundos el proceso de revisión de las pruebas analíticas de sangre de los pacientes, la app Hark estaba configurada para reducir las listas de espera y el papeleo en los hospitales. Al inicio de esta colaboración de Google con el NHS, los responsables de DeepMind Health se comprometieron a que los datos de los pacientes (registros personales de 1,6 millones de pacientes) nunca se conectarían con las cuentas o servicios de Google, precisamente en aras del respeto a su derecho a la intimidad (Powles-Hodson, 2017, 351-367).

En 2019 Google rompería esta promesa inicial al vincular DeepMind Health, la empresa filial de atención médica a Google, la empresa matriz, incorporando a la misma millones de datos personales de los pacientes del NHS sin contar con su previo consentimiento, a la vez que desmantelaba el comité de revisión

ética de los algoritmos aplicados en el ámbito sanitario y cuyo panel de revisores ya había manifestado pocos meses antes del cierre del mismo su preocupación ante el riesgo potencial de que DeepMind Health pudiera utilizar su acceso a los datos personales de los pacientes del NHS para propiciar las ganancias monopolísticas para la empresa matriz de Google, Alphabet (Murgia, 2018).

A propósito de la revisión institucional de la transparencia y buen funcionamiento de los algoritmos, hay autores que ponen en duda que un solo comité ético sea capaz por sí mismo de garantizar el rigor, la utilidad y la integridad de los *big data*. Se trataría, según esta doctrina escéptica respecto a la viabilidad del control de humanidad sobre la IA, de una pretensión poco realista (Lipworth-Mason-Kerridge-Ioannidis, 2017, 489-500). Otra línea doctrinal sostiene que el problema de la responsabilidad y de la pública rendición de cuentas por parte de quienes hacen uso de los algoritmos potencialmente lesivos de derechos y libertades de los ciudadanos debería ser planteada como una cuestión de acuerdo o consenso entre las partes interesadas (Floridi, 2022, 173).

Como solución alternativa a la posible falta de acuerdo sobre la revisión de la transparencia y el buen funcionamiento de los algoritmos, existe un código ético propuesto por la *Association for Computing Machinery* (ACM) en el que se enuncian siete principios éticos fundamentales para los profesionales de la Informática y la Computación⁴. El objetivo principal que se

⁴ Los siete principios éticos generales enunciados por el Código de Ética y Conducta Profesional de la ACM y que todo buen profesional de la Informática y la Computación debería seguir son los siguientes: 1.- Contribuir a la sociedad y al bienestar humano, reconociendo que todas las personas son partes interesadas en la Informática; 2.- Evitar daños o consecuencias negativas, especialmente cuando son significativas e injustas; 3.- La conducta de un buen profesional debe ser siempre transparente, honesta y fiable; 4.- Respeto a los valores de igualdad (no discriminación), tolerancia, respeto a los demás y justicia; 5.- Crear oportunidades para que los miembros de la organización o grupo crezcan como profesionales; 6.- Respetar la intimidad (*privacy*), por lo que los profesionales de

persigue este código deontológico profesional es, precisamente, el beneficio de todas las partes interesadas en el mantenimiento de un debate abierto entorno a las cuestiones éticas que promueven la responsabilidad y la transparencia (Buhmann-Passmann-Fieseler, 2019).

Por último, entre quienes apuestan por generar las condiciones necesarias para el cultivo de una “cultura algorítmica” más responsable, destaca Luciano Floridi, quien propone repartir una “responsabilidad moral distributiva” entre todos los agentes morales, es decir, personas humanas o jurídicas, es decir, sociedades constituidas por seres humanos (Floridi, 2016, 2). En sentido análogo, Mark Coeckelbergh, inspirándose en la Declaración de Montreal para un desarrollo responsable de la IA (2018), sostiene que la espiral de la Ética de la IA no debe circunscribirse solamente al ámbito de la política internacional y la gobernanza global; en efecto, también la comunidad académica debe asumir una posición activa en la propuesta de iniciativas que busquen una interacción equilibrada e inocua entre los humanos y las máquinas desarrolladas con IA, de tal modo que sea posible concebir la armonía entre el avance tecnológico con la salvaguarda del legado humanista y los valores que son inherentes y necesarios para la defensa de la dignidad, la autonomía de la voluntad y los derechos fundamentales del ser humano (Coeckelbergh, 2020, 157).

BIBLIOGRAFÍA

BALASUBRAMANIAN, R.-LIBARIKIAN, A.-McELHANEY, D. (2021), “Insurance 2030-The Impact of AI on the future of insurance”, *McKinsey*

la Informática solo deben usar la información personal para fines legítimos y sin violar los derechos y libertades de los individuos; 7.- Respeto de la confidencialidad de los secretos comerciales, los datos de los clientes, las estrategias comerciales no públicas, información financiera, datos de la investigación, artículos académicos previos a la publicación y solicitudes de patentes. Cfr., [https:// www.acm.org/code-of-ethics](https://www.acm.org/code-of-ethics)

& Company, March 12, 2021,
<https://www.mckinsey.com/industries/financial-services/our-insights/insurance-2030-the-impact-of-ai-on-the-future-of-insurance>
 (última consulta, 25 de abril de 2023)

BENON, H.-DECKER, M. (2021), "Insuring Commercial Drones? Liability or Opportunity? *Insurance Journal*, September 6, 2021, <https://www.insurancejournal.com/magazines/mag-features/2021/09/06/630181.htm> (última consulta, 25 de abril de 2023)

BUHMANN, A.-PASSMANN, J.-FIESELER, C. (2019), "Managing algorithmic accountability: Balancing reputational concerns, engagement strategies, and the potential of rational discourse", *Journal of Business Ethics*, nº 163, pp. 265-280, <https://link.springer.com/article/10.1007/s10551-019-04226-4> (última consulta, 25 de abril de 2023)

BOIX PALOP, A. (2022), "Transparencia en la utilización de Inteligencia Artificial por parte de la Administración", *El Cronista del Estado social y democrático de derecho*, nº 100, pp. 90-105.

CERRILLO I MARTÍNEZ, A. (2021), "La transparencia de los algoritmos que utilizan las administraciones públicas", *Anuario de Transparencia Local 3/2020*, Fundación Democracia y Gobierno Local, pp. 41-78.

COECKELBERGH, M. (2020), *AI Ethics*, The MIT Press, Cambridge (Massachusetts)-London.

FITZEK, F. H. P.-LI, S.-Ch.-SPEIDEL, S.-STRUFE, T. (2021), "Tactile Internet with Human-in-the-Loop: New Frontiers of Transdisciplinary Research", *Tactile Internet with Human-in-the-Loop*, Academic Press, San Diego (California).

FLORIDI, L. (2016), "Faultless responsibility: On the nature and allocation of moral responsibility for distributed moral actions", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374 (2083), <https://royalsocietypublishing.org/doi/10.1098/rsta.2016.0112> (última consulta, 25 de abril de 2023)

(2022), *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, Raffaello Cortina Editore, Milano.

GARZÓN VALDÉS, E. (1989a), "Representación y democracia", *Doxa. Cuadernos de Filosofía del derecho*, nº 6, 1989, pp. 143-164.

(1989b), "Algo más acerca del coto vedado", *Doxa. Cuadernos de Filosofía del derecho*, nº 6, 1989, pp. 209-213.

GROTE, T.-BERENS, P. (2019), "On the Ethics of Algorithmic Decision-Making in Healthcare", *Journal of Medical Ethics*, nº 46, 20th November

2019, pp. 205-211, [https:// 10.1136/medethics-2019-105586](https://10.1136/medethics-2019-105586) (última consulta, 25 de abril de 2023)

HABERMAS, J. (1981), *Theorie des Kommunikativen Handelns* (2 Bände), Suhrkamp, Frankfurt am Main.

HERBERT, B.-ANDERSON, K. J. (2011⁵), *Dune. La Yihad butleriana*, trad. esp., Eduardo G. Murillo, Debolsillo, Barcelona.

HUERGO LORA, A. (2020), “Una aproximación a los algoritmos desde el Derecho administrativo” (pp. 23-87), en *La regulación de los algoritmos* (Director: Alejandro Huergo Lora; coordinador: Gustavo Manuel Díaz González), Thomson Reuters Aranzadi, Cizur Menor (Navarra).

JIE, Z.-ZHIYING, Z.-LI, L. (2021), “A Meta-Analysis of Watson for Oncology in Clinical Application”, *Nature*, 11:5792, <https://www.nature.com/articles/s41598-021-84973-5> (última consulta, 25 de abril de 2023)

LARSON, E. J. (2022), *El mito de la Inteligencia Artificial. Por qué las máquinas no pueden pensar como nosotros lo hacemos*, trad. esp., Milo J. Krmpotić, Shackleton Books, Barcelona.

LIPWORTH, W.-MASON, P. H.-KERRIDGE, I.-IOANNIDIS, J. P. A. (2017), “Ethics and Epistemology in Big Data Research”, *Journal of Bioethical Inquiry*, 14 (4), pp. 489-500, <https://link.springer.com/article/10.1007/s11673-017-9771-3>

LOSANO, M. G. (1969), *Giuscibernetica. Macchine e modelli cibernetici del diritto*, Einaudi, Torino.

MEDINA GUERRERO, M. (2022), “El derecho a conocer los algoritmos utilizados en la toma de decisiones. Aproximación desde la perspectiva del derecho fundamental a la protección de datos personales”, *Teoría y realidad constitucional*, nº 49,

NINO, C. S. (1989), “Autonomía y necesidades básicas”, *Doxa. Cuadernos de Filosofía del derecho*, nº 6, 1989, pp. 21-34.

MURGIA, M. (2018), “DeepMind’s move to transfer health unit to Google stirs data fears”, *Financial Times*, 13th November, <https://www.ft.com/content/f4a73450-e771-11e8-8a85-04b8afea6ea3> (última consulta, 25 de abril de 2023)

NUSSBAUM, M. C. *Frontiers of Justice. Disability, Nationality. Species Membership* (2007), The Belknap Press of Harvard University Press, Cambridge (Massachusetts)-London.

(2012), *Crear capacidades. Propuesta para el desarrollo humano*, trad. esp., A. Santos Mosquera, Paidós, Barcelona.

PONCE SOLÉ, J. (2019a), *La lucha por el derecho a una buena administración: el estándar jurídico de diligencia debida y el buen*

gobierno en las políticas públicas, Cuadernos de la Cátedra de Democracia y Derechos Humanos, núm. 15, Universidad de Alcalá de Henares- Defensor del Pueblo, Alcalá de Henares.

(2019b), "Inteligencia artificial, Derecho administrativo y reserva de humanidad, algoritmos y procedimiento administrativo debido tecnológico", *Revista General de Derecho Administrativo*, nº 50, pp. 141-171.

(2022), "Reserva de humanidad y supervisión humana de la Inteligencia Artificial", *El Cronista del Estado social y democrático de derecho*, nº 100, pp. 58-67.

POWLES, J.-HODSON, H. (2017), "Google DeepMind and Healthcare in Age of Algorithms", *Health and Technology*, 7/2017, pp. 351-367, <https://link.springer.com/article/10.1007/s12553-017-0179-1>

PRESNO LINERA, M. (2022), "Derechos fundamentales e Inteligencia Artificial en el Estado social, democrático y digital de Derecho", *El Cronista del Estado social y democrático de derecho*, nº 100, pp. 48-57.

RAO, D. M.- CHERNYAKHOVSKY, A.-RAO, V. (2011), "Analyzing Global Epidemiology of Diseases Using Human-in-the-Loop Bio-Simulations" (153-174), *Human-in-the-Loop Simulations* (eds. L. Rothrock-S. Narayanan), Springer Verlag, London.

RAWLS, J. (1971), *A Theory of Justice*, The Belknap Press of Harvard University Press, Cambridge (Massachusetts)-London.

ROSS, M.-TAYLOR, J. (2021), "Managing AI Decision-Making Tools", *Harvard Business Review*, <https://hbr.org/2021/11/managing-ai-decision-making-tools> (última consulta, 25 de abril de 2023)

SHACKLE, S. (2020), "The Mystery of the Gatwick Drone", *The Guardian*, 1 December 2020, <https://www.theguardian.com/uk-news/2020/dec/01/the-mystery-of-the-gatwick-drone> (última consulta, 25 de abril de 2023)

ŽIŽEK, Slavoj (2023), *Hegel y el cerebro conectado*, trad. esp., Fernando Borrajo, Paidós, Barcelona.

¿SE PUEDE AUTOMATIZAR LA *FAIRNESS* EN LA INTELIGENCIA ARTIFICIAL? LA PROBLEMÁTICA DE LA DISCRIMINACIÓN Y LA DIMENSIÓN DE LA INTERSECCIONALIDAD

Nuria Beloso Martín
Catedrática de Filosofía del Derecho
Universidad de Burgos (España)

INTRODUCCIÓN

Los usos de la IA en el Derecho han supuesto numerosos retos que acaban confluyendo en la necesidad de que la justicia presida el diseño, el funcionamiento y las decisiones que se adoptan a partir del algoritmo. La proyección de la equidad en el Derecho antidiscriminatorio y en las políticas de igualdad, que ya en el ámbito analógico no es pacífica, con mayor razón presenta particulares dificultades cuando se usa la inteligencia artificial (en adelante, IA). Dejando a un lado la desigualdad en el acceso (“brecha digital”), pretendo centrarme en el examen de si se puede automatizar la *fairness* en la inteligencia artificial.

La justicia requiere la aplicación imparcial en unos casos, pero en otros, como ya advirtió Aristóteles, la justicia también implica la idea de trato proporcional, lo que implica que los destinatarios obtengan cantidades desiguales de cualquier bien en cuestión (Aristóteles, cap. 3). La igualdad aquí es el principio por defecto que se aplica en ausencia de reivindicaciones especiales que puedan presentarse como razones de justicia. La dificultad radica en identificar qué sea ese '*sum cuique*', lo que le corresponda a cada uno -sea una persona individual o se trate de un grupo con determinados atributos-. Hay varias categorías que se convierten

en imprescindibles cuando la filosofía y la *Machine Learning* (ML) van de la mano, y que permiten entender las dificultades y limitaciones de *la construcción de un modelo de predicción y de decisión perfecto: parcialidad, opresión, igualdad y discriminación.*

La búsqueda de la igualdad y de la no-discriminación, por la que tanto se trabaja en el ámbito *offline*, se traslada, y con mayor intensidad, al ámbito digital y de la IA. Las desigualdades de género, raza, sexo, clase social, etnia, religión, -entre otras- están presentes en los sistemas de clasificación de IA, en los sistemas predictivos y en los sistemas decisionales, amplificando y perpetuando tales desigualdades. Por ello, hay que insistir en iniciativas que delinee métodos y procedimientos para identificar, aminorar y corregir esos sesgos en los sistemas de IA. Cuando el entrenamiento de aprendizaje automático incide y afecta a atributos sensibles relacionados con las personas, la consecución de la equidad y de la imparcialidad (decisión justa) resulta dificultosa. Esta misma consideración de atributos sensibles debe desarrollarse en la formulación de los algoritmos, en unos casos, para establecer tales discriminaciones positivas y, en otros, para eliminar diferencias.

Las nociones de *fairness* propuestas por la comunidad de aprendizaje automático se hacen eco de ciertas consideraciones de justicia y han dirigido su mirada a filósofos morales y políticos. Varios trabajos recientes han señalado la necesidad de reflexionar sobre tales conexiones de manera que técnicos, juristas, filósofos y éticos trabajen conjuntamente (Barocas et al., 2020; Verma; Rubin, 2018).

En aras de evitar malentendidos conceptuales, hago la advertencia preliminar de que justicia y equidad no son términos sinónimos (Goldman y Cropanzano, 2015), al igual que tampoco lo son, a su vez, equidad e imparcialidad. Precisar tales conceptos requeriría un estudio profundo que ahora excedería los límites propuestos para este trabajo. A riesgo de simplificar en exceso, considero que

la justicia, en su conceptualización de valor y fin del Derecho, sería la categoría omnicomprensiva tanto de la equidad como de la imparcialidad, entendiendo por estos criterios instrumentales el enfoque dirigido a lograr el fin más amplio que es la justicia (Belloso Martín, 2023). En este trabajo los utilizaré indistintamente.

La mayoría de las investigaciones sobre sesgos algorítmicos, se suelen enfocar en una sola variable de análisis cada vez, bien sea el sexo o la raza, pero no realizan una valoración sobre varios ejes de desigualdad y que actúen conjuntamente. Ello dificulta detectar y conocer las consecuencias para los sujetos que pertenecen a más de uno de los grupos protegidos¹. Esta dificultad se pone especialmente de relieve cuando se relaciona con la justicia distributiva y con los derechos económicos, sociales y culturales, es decir, en cómo establecer ese *suum cuique tribuere* automatizado. Asimismo, si un procedimiento algorítmico, bien sea predictivo o decisonal, ya resulta técnica y jurídicamente complejo para que se ajuste al *fairness*, mayor dificultad se presenta cuando debe lidiar con una dimensión de interseccionalidad.

En este trabajo me propongo analizar si se puede automatizar la *fairness* de la IA, para lo que examinaré algunas de las dificultades que debe superar el diseño de un algoritmo -siempre desde el enfoque de la filosofía jurídica- para que se ajuste al *fairness* (bien se entienda como equidad o como imparcialidad). Entre esas dificultades revisaré principalmente dos: la dimensión de interseccionalidad como factor que potencia la discriminación; y la dificultad para fijar un estándar homogéneo y estable que sirva como referencia para que un proceso decisonal de una IA establezca si se ha producido o no discriminación, a diferencia del

¹ Como explica Beltramelli et al., son varios los especialistas que han puesto de manifiesto los problemas de realizar análisis en un solo eje de este campo. Así, Buolamwini recomienda tomar como punto de partida el marco interpretativo de las estructuras de dominación y desigualdad, que propone la interseccionalidad, para analizar el desarrollo desigual en las intersecciones de las variables género y etnia-raza en algoritmos de clasificación a partir del reconocimiento facial.

criterio judicial en el contexto de la Unión Europea, que trabaja con parámetros contextuales.

Reducir o minimizar sesgos, velar por el respeto al principio de igualdad de oportunidades, neutralizar prejuicios, resolver la dicotomía discriminación directa/discriminación indirecta, son sólo algunas de las operaciones que conlleva, en la vida analógica, lidiar con la discriminación. En el ámbito de la IA, tanto la problemática que conlleva la discriminación como la dimensión de interseccionalidad (necesidad de clasificar, distinguir grupos, subgrupos, atributos sensibles, ventajas merecidas, etc.) ponen de manifiesto la dificultad de englobar tales operaciones de calibrado en un estándar fijo y homogéneo en las predicciones y decisiones automatizadas.

Los interrogantes que se suscitan son numerosos: ¿qué es la justicia interseccional? ¿Qué relación guarda con la justicia individual y grupal? ¿puede hablarse de una justicia en la ciencia de datos? ¿qué variables se deben de tomar en consideración para el diseño de un algoritmo sensible a la justicia interseccional? ¿qué criterios habría que aplicar para conseguir calibrar la *fairness* de las decisiones algorítmicas? ¿se podrían diseñar unos estándares que permitieran establecer unos umbrales para determinar la *fairness* del algoritmo? Un examen exhaustivo de todas estas cuestiones excedería los límites que me he propuesto en este trabajo. Por ello, voy a limitarme a ofrecer algunas claves sobre la temática.

A VUELTAS CON LA IGUALDAD Y NO-DISCRIMINACIÓN

Una de las proyecciones de la justicia es la igualdad y no-discriminación. Como bien explica Castilla, de manera simplificada, la igualdad ante la ley significa que las situaciones iguales deben ser tratadas iguales y las situaciones desiguales deben ser tratadas desigualmente, siempre de manera justificada objetiva, razonable y proporcional. Por su parte, la prohibición de discriminación

constituye una garantía reforzada del principio de igualdad. La discriminación no solo incluye toda distinción, sino también toda exclusión, restricción o preferencia. Si cualquiera de esas cuatro acciones citadas se hace por motivos de la raza, el color, el sexo, el idioma, la religión, la opinión política o de otra índole, el origen nacional, étnico o social, la posición económica, los impedimentos físicos, el nacimiento o cualquier otra condición, de cualquier persona, se podría estar discriminando (2022:16).

La doctrina de la no discriminación se ha ido construyendo a partir de varias clasificaciones, tales como la distinción entre discriminación directa e indirecta; la discriminación múltiple -que no debe confundirse con la discriminación interseccional-; y la discriminación estructural o sistémica. Estas clasificaciones no sólo tienen importantes repercusiones en la vida analógica para el ámbito jurídico sino también en su aplicación en la IA.²

² Por ello, y siguiendo a Castilla, conviene recordar aquí en qué consiste cada tipo de discriminación.

“La discriminación es directa cuando los actos, leyes o políticas públicas desventajosas están dirigidas expresamente a una categoría, grupo o persona. Es indirecta cuando esos actos, leyes o políticas públicas aplican un criterio aparentemente neutral al no dirigirse a una categoría, grupo social o persona, pero impone a estos una carga desproporcionada o desventaja particular con respecto a otras personas que les perjudica o imposibilita colocarse en el supuesto previsto”.

La discriminación múltiple no debe confundirse con la interseccionalidad, ya que, aunque son muy parecidas, en la primera se dan discriminaciones por diversas razones en diferentes o un mismo momento. Es decir, la discriminación múltiple, refleja la situación en la que una persona, por sus características personales, pertenece a varios grupos que son objeto de discriminación y sufre diferentes discriminaciones en diferentes momentos o en un mismo momento, pero por varias razones. En cambio, la interseccionalidad “es una metodología que sirve para analizar la forma en la que varios tipos de desigualdad y discriminación inciden y operan de manera conjunta en una persona por las características personales que esta tiene, así como para establecer la forma en la que dichas causas de desigualdad y discriminación se exacerban mutuamente y multiplican entre sí de acuerdo al contexto social específico en el que son puestas de manifiesto o identificadas” (*Derechos humanos desde una perspectiva*

LA DIMENSIÓN DE INTERSECCIONALIDAD: SU RECONOCIMIENTO JUDICIAL

Hay estados opresivos que se entrecruzan y que influyen en el tratamiento legal, para lo que se recurre al Derecho y, sobre todo, a la justicia y a la equidad, para buscar soluciones y metodologías que puedan ser utilizadas para corregir las inequidades. Hay categorías sociales estables y duraderas como la raza, el sexo y la clase. Los investigadores pueden seleccionar una categoría amplia - como las mujeres - y luego tratar de identificar las diferencias que se producen entre las mujeres a partir de otras categorías sociales (raza, por ejemplo). Hay grupos en los que emergen intersecciones específicas dando lugar a mujeres, negras, pobres, o bien, mujer, indígena, con necesidades especiales, pobre, y de avanzada edad. La interseccionalidad contribuye a poner en entredicho los análisis que pretenden reducir a los sujetos de estudio a variables estadísticas e independientes a la vez que ofrece una nueva dimensión de búsqueda de solución a los problemas que trascienden los razonamientos y estructuras binarias que son las que priman en este campo.

Crenshaw, jurista norteamericana, publicó en 1989 un trabajo que ha sido una referencia en el tema, en el que se propuso poner en cuestión “la tendencia a tratar la raza y el género como categorías de experiencia y análisis mutuamente excluyentes”. Crenshaw se dirigió a demostrar dos cuestiones. La primera, que la mera yuxtaposición de los ejes de análisis hace que las mujeres negras

interseccional, Institut de Drets Humans de Catalunya, Barcelona, 2021. <https://www.idhc.org/arxius/recerca/1633338099-InterseccionalidadyDDHH.pdf>.

Por último, se podría hablar de la discriminación estructural o sistémica, esto es, del conjunto de normas, reglas, rutinas, patrones, actitudes y pautas de comportamiento, tanto de jure como de facto, que dan paso a una situación de inferioridad y exclusión contra un grupo de personas de forma generalizada, las cuales son perpetuadas a lo largo del tiempo e incluso por generaciones (Cfr. Comisión Interamericana de Derechos Humanos, Informe sobre pobreza y derechos humanos en las Américas, 2017, párr. 393.

<http://www.oas.org/es/cidh/informes/pdfs/pobrezaddhh2017.pdf>).

desaparezcan a nivel teórico; y la segunda, que las limitaciones que conlleva ese planteamiento separador destruyeran cualquier intento de ampliar los análisis feministas y antirracistas. Es decir, con respecto a la primera cuestión, denuncia que, en los casos de discriminación racial, tal discriminación se visualiza en varones negros (por tanto, privilegiados por su sexo o, eventualmente, clase), mientras que, en los casos de discriminación sexual, la discriminación se focaliza en mujeres blancas (es decir, privilegiadas por su raza y, eventualmente, clase). Con respecto a la segunda cuestión, señala que a veces las mujeres negras son excluidas de la teoría feminista, por un lado, y de la política antirracista, por otra, ya que ambos discursos se levantan sobre unas experiencias que no reflejan la interacción de género y raza. Crenshaw (1991) argumenta que las personas pueden experimentar la opresión de manera diferente en función de su superposición identidades, a lo que denomina como “interseccionalidades”. Por ejemplo, una mujer de raza negra y pobre puede experimentar opresión no experimentada por otras mujeres, o por otras personas de raza negra o por otras personas que dispongan de una economía saneada.³

Advierte de las consecuencias de tal construcción a nivel jurídico y judicial, lo cual se refleja en uno de los casos que expone, al que voy a referirme seguidamente, en cuanto permitirá entender mejor

³ Las estructuras sociales son poderosos agentes de desigualdad, discriminación, prejuicios y parcialidad. Como ha apuntado Young, las personas ocupan diferentes posiciones dentro de una “sociedad estructurada”, cada una de ellas con sus correspondientes expectativas, ventajas, y desventajas. Las estructuras sociales se convierten en lugares de injusticia cuando empoderan sistemáticamente a las personas en algunas posiciones al desempoderar a otros (Young, 2000) -como en el caso de los hombres en el mercado laboral cuando. Un ejemplo es el del algoritmo que utilizaba Amazon para la contratación laboral, y que rechazaba los currículos de las mujeres porque había sido entrenado a partir de la contratación de hombres, lo cuales eran empoderados porque las mujeres estaban desempoderadas-. Un algoritmo puede aplicar formas de acción positiva o acciones afirmativas para compensar de manera directa algunas categorías desaventajadas, especialmente, algún grupo desventajado, actuando de este modo como una forma de “acción positiva”.

qué se la interseccionalidad. Cinco mujeres negras demandaron a la empresa multinacional General Motors alegando que el sistema de antigüedad de la empresa venía a perpetuar los efectos de la discriminación pasada hacia las mujeres negras. Su demanda se basaba en que las mujeres negras habían sido contratadas por la empresa a partir de 1970 y, por lo tanto, más tarde que las mujeres blancas (que lo habían sido antes de 1964), pero también más tarde que los hombres negros (que lo habían sido después de 1964 pero con anterioridad a 1970). Como consecuencia de esta cronología en las contrataciones, al llegar una fase de recesión, la empresa comenzó a despedir al personal ajustándose el criterio de la antigüedad, de manera que las mujeres negras fueron las primeras en ser despedidas. El tribunal justificó que no aceptó la demanda porque en el Título VII de la ley no está prevista la posibilidad de combinar raza y género para fundar la protección antidiscriminatoria. Es decir, el tribunal razonaba que, puesto dado que la General Motors había contratado a mujeres (aunque fueran blancas) antes de 1964, no podía haber discriminación por sexo que el criterio de antigüedad pudiera perpetuar; y dado que, asimismo, había contratado hombres negros, tampoco discriminación racial (Barrere Unzueta, 2010: 231).⁴

A juicio de Crenshaw, el tribunal lleva a cabo una doble-discriminación (*double discrimination*) puesto que se ha producido un proceso injusto de identificación de una clase o categoría

⁴ Como acertadamente advierte Barrere Unzueta, el hecho de que se subraye este aspecto positivo de la interseccionalidad no implica que ciertas exralimitaciones de la misma conduzcan a resultados adversos, como, por ejemplo, si tal concepto origina un proceso infinito de división de grupos sociales en el cual, al final, no quedan más que individuos. Es decir, una versión extrema de la interseccionalidad disolvería o haría insignificante la utilización de términos como “clase”, “hombres negros”, “mujeres” o “mujeres negras”, en tanto tales grupos se encontrarían fracturados por otras estructuras sociales. Por ello, conviene recordar que “la idea de la interseccionalidad surgió de que las estructuras sociales se entrecruzan para dar forma a posiciones sociales” (*op.cit.*, p.250). En definitiva, “la perspectiva interseccional de la discriminación obliga a una revisión del Derecho antidiscriminatorio clásico basado en la yuxtaposición de los factores de discriminación” (*ibidem*, p.251).

específica, bien a través de la intencionalidad discriminatoria que identifique tal categoría, o bien mediante la demostración de las desventajas de todos los miembros de esa categoría.

La condición de pobreza, la condición de mujer, y otros atributos sensibles, son los que los altos tribunales han tomado en consideración para construir la dimensión de interseccionalidad. Como explica Castilla, desde el ámbito judicial, la Corte Interamericana de Derechos Humanos utilizó por primera vez el concepto de “interseccionalidad” en el análisis de la discriminación sufrida por una niña en el acceso a educación en el caso *Gonzales Lluy y Otros Vs. Ecuador* (2015). Si bien la sentencia no desarrolla el concepto, el párrafo 290 de esa sentencia explica que:

confluyeron en forma interseccional múltiples factores de vulnerabilidad y riesgo de discriminación asociados a su condición de niña, mujer, persona en situación de pobreza y persona con VIH. La discriminación que vivió Talía no sólo fue ocasionada por múltiples factores, sino que derivó en una forma específica de discriminación que resultó de la intersección de dichos factores, es decir, si alguno de dichos factores no hubiese existido, la discriminación habría tenido una naturaleza diferente. En efecto, la pobreza impactó en el acceso inicial a una atención en salud que no fue de calidad y que, por el contrario, generó el contagio con VIH. La situación de pobreza impactó también en las dificultades para encontrar un mejor acceso al sistema educativo y tener una vivienda digna.

En el ámbito del Tribunal Europeo de Derechos Humanos, Castillo señala que el caso *B.S. Vs. España* (2012) es considerado como el primero en el que se utiliza una “perspectiva interseccional”, aunque no se mencione expresamente el término en la sentencia, ni tampoco se ofrezca un análisis especial sobre el tema, -a pesar de que algunas partes intervinientes en el caso así lo solicitaron expresamente al Tribunal, al señalarle la existencia de

“discriminación multifactorial”-. No obstante, en el párrafo 71 de la sentencia, el Tribunal Europeo reconoce que las autoridades españolas “no tuvieron en cuenta la vulnerabilidad específica de la demandante, inherente a su condición de mujer africana ejerciendo la prostitución.”

La dimensión de interseccionalidad del mundo *offline* va acompañada de una visibilidad que la hace más fácilmente detectable y, por tanto, susceptible de neutralizar o de aminorar. Distinta es la situación en el universo algorítmico, en el que su visibilidad es bastante más escasa, lo que hace necesario un análisis del estado de la cuestión, al que me referiré seguidamente.

Resulta necesario abordar la interacción entre distintas desigualdades a la hora de diseñar e implementar las políticas públicas orientadas a un enfoque de igualdad. Sin embargo, hay también que tomar en consideración que, si se llevan hasta el extremo, se puede acabar desembocando en una tiranía de las minorías. La reserva de cupos para individuos con ciertos atributos puede derivar en minusvalorar su actividad y competencia al considerar que, por ejemplo, se les contrata para una determinada función por el hecho de aunar una o varias desigualdades, y no por su particular cualificación. La ideología que subyace a estas políticas hace que la cuestión no sea pacífica.

LA JUSTICIA INTERSECCIONAL ALGORÍTMICA: LA DIFFERENTIAL FAIRNESS

Los algoritmos de aprendizaje automático supervisados son, de por sí, de naturaleza discriminativa (Stewart, 2020). La razón de tal discriminación estriba en que muchos de los algoritmos se han diseñado precisamente para clasificar los datos en función de las informaciones que aparecen incrustadas en los mismos. A diferencia del "algoritmo de generación" que genera datos según

una categoría específica, el “algoritmo discriminatorio” clasifica, divide los datos en diversas categorías, conforme a los criterios que el programador haya establecido, y realiza distribuciones. Los algoritmos son modelados para reconocer características y atributos que en situaciones anteriores han conducido al éxito, no interesándose por lo justo sino replicando modelos y prácticas anteriores.

La identificación de sesgos y su neutralización es uno de los más importantes desafíos que tiene ante sí la IA. Con relación a este reto, las carencias en la detección de sesgos resultan aún más manifiestas cuando se toman variables de forma independiente, ya que las desigualdades pueden presentarse en la intersección de las mismas y, en cambio, de forma individual, resultan indetectables. Es decir, cuando se cruzan distintas categorías que han forjado, a lo largo de la historia, situaciones de discriminación, de opresión y de diferencia injusta, además de hacer más compleja la detección de los sesgos, también incrementa y potencia tal discriminación. No se trata de considerar negativa la diferencia sino la diferencia injusta o no correcta.

Uno de los grandes retos para los juristas -y también para los técnicos- es el de indagar si es posible un algoritmo equitativo, un sistema de IA que incorpore *fairness*. *Fairness* puede definirse como “la ausencia de perjuicio o preferencia por un individuo o grupo basado en sus características” (Ghosh; Genuit; Reagan, 2021: 2) por parte de un sistema cuando proceda a realizar una tarea de clasificación, predicción y toma de decisiones basada en datos.

Una de las definiciones de justicia interseccional más citadas por parte de quien investiga sobre esta temática es la de Foulds et al. (2019), que desarrollan el concepto de *differential fairness*. A juicio de los autores, la definición debe satisfacer los siguientes criterios: deben considerarse múltiples atributos protegidos; todos los

valores de intersección de los atributos protegidos deberían estar protegidos; se debe garantizar la protección de los valores individuales de los atributos protegidos (por ejemplo “las mujeres”, en el caso de “mujeres afro”); la definición debe proteger a los grupos minoritarios; y la definición debería garantizar que las diferencias sistemáticas entre los grupos protegidos, se rectifiquen, en lugar de codificarse (Foulds et al., 2019). Proponen la creación de subgrupos a partir de la intersección de dos o más variables protegidas, garantizando la paridad estadística entre tales subgrupos. Así, un algoritmo de IA es justo si las probabilidades de los resultados (por ejemplo, ser aprobado/a para un préstamo por parte de una entidad bancaria) son iguales o similares, independientemente de la combinación de intersección de atributos de un grupo, como el género y la raza.⁵

¿SE PUEDE AUTOMATIZAR LA EQUIDAD (FAIRNESS)?

En el ámbito norteamericano se pueden encontrar numerosos estudios sobre el sesgo, la discriminación y *fairness* en la inteligencia artificial y en el aprendizaje automático. Sin embargo, en el ámbito de la Unión Europea son escasos, lo cual contrasta con la relevancia que en el contexto jurídico y judicial europeo se ha dado al derecho antidiscriminatorio. Junto a los escasos estudios dedicados a *fairness* algorítmica, también se ha señalado que hay una brecha entre las herramientas estadísticas de *fairness*

⁵ En la literatura se encuentran una diversidad de “casos” -unos se basan en sistemas de clasificación, y otros se basan en sistemas de predicción- que permiten visualizar ejemplos para dar cuenta de los sesgos producidos al no tomar en cuenta las diferentes dimensiones que atraviesan a las personas. Son ya conocidos que diversos datos comerciales utilizados para entrenar algoritmos de reconocimiento facial, a partir del muestreo de cómo varios sistemas de clasificación de imágenes faciales por género van actuando, reflejan que la calibración de los sistemas para las mujeres de piel oscura es un 30% menos preciso en comparación al que se realizaba en hombres de piel clara.

algorítmica y las métricas de discriminación que, como requisitos probatorios, utilizan los tribunales europeos.

Wachter et al. (2021) señalan que, si en un caso de discriminación, se establece una comparativa entre un proceso decisorial desarrollado por un tribunal europeo, frente a una decisión de una IA, habría claras diferencias. Muchos de los conceptos fundamentales para presentar una demanda, como la composición del grupo desfavorecido y el favorecido, la gravedad y el tipo de daño sufrido, los requisitos para la pertinencia y admisibilidad de las pruebas, requieren que el Poder judicial tome decisiones normativas o políticas sobre una base de caso por caso, de manera que los requisitos actuales en el marco de la Unión Europea (UE) son demasiado contextuales, dependen de la intuición y están abiertos a la interpretación judicial. Es decir, caso por caso se valora si la magnitud de la discriminación depende en gran medida del tipo de daño cometido, si un grupo minoritario se ve afectado, y si la discriminación refleja una injusticia sistémica medida contra el marco legal. Wachter et al. engloban este proceder en lo que denominan “igualdad contextual”.

A partir de esta caracterización del sistema judicial en casos de discriminación, consideran que en la Unión Europea no se puede automatizar la *fairness* o la no discriminación porque el propio diseño legal no proporciona un marco estático u homogéneo adecuado que permita probar la discriminación en los sistemas automatizados de una IA. Por ejemplo, la decisión de una Administración pública en orden a conceder una ayuda social,⁶ o

⁶ En España se puede citar el caso del Programa BOSCO (Bono Social Eléctrico). Tal Bono consiste en un descuento en la factura eléctrica para los "consumidores vulnerables y con menor capacidad económica para afrontar los precios elevados". La norma establece tres vías de entrada para acceder al descuento: i) por rentas bajas; ii) familias numerosas; iii) 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 -es decir, no tuvo en cuenta la dimensión de interseccionalidad (mujer, escasos recursos, viuda). En el caso de una viuda, la

de una empresa a la hora de contratar, será o no susceptible de ser considerada como discriminatoria, dependiendo en buena medida de la valoración e interpretación de la legislación y del contexto que, en cada caso, hago el juez. Es como si la horquilla que mide el umbral de discriminación para probar que una determinada decisión es discriminatoria, fuera flexible. Si lo que se pretende es la automatización de este tipo de decisiones, la falta de unos estándares fijos y homogéneos, hacen muy difícil la consecución de una decisión algorítmica ajustada a *fairness*.⁷

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 (septiembre de 2019), interpuso recurso contencioso-administrativo 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. En el mes de febrero de 2022, se dictó Sentencia desestimatoria para Civio, y con una condena en costas de 2.000 €. Según la sentencia, sí, es un acto administrativo pero, a diferencia de lo que interpreta Civio, quien decide no es la aplicación, sino un órgano administrativo.

⁷ Traigo aquí a colación la reciente polémica que ha surgido en España con respecto a la concesión del bono social térmico. El Bono Social Térmico (BST) es una **ayuda directa a consumidores vulnerables destinada a compensar los gastos de suministro de energía para los usos térmicos**, como calefacción, agua caliente sanitaria o cocina. Su regulación se establece en el Real Decreto-ley 15/2018, de 5 de octubre, de medidas urgentes para la transición energética y la protección de los consumidores, con el fin de complementar la ayuda percibida en concepto de Bono Social Eléctrico por los “consumidores vulnerables”, para la energía destinada a calefacción, agua caliente sanitaria o cocina, independientemente de cuál sea la fuente utilizada. Es decir, son beneficiarios de este Bono, **sin necesidad de realizar ningún trámite ni solicitud**, los beneficiarios del Bono Social Eléctrico. A tenor de la denominación de estas normativas (Bono eléctrico de justicia energética del Real Decreto-ley 18/2022, de 18 de octubre.) se deduce que estas ayudas constituyen una proyección de la justicia (justicia energética” la denomina el legislador), y que se articulan a través de programas informáticos específicos. La legislación contempla que el colectivo de familias numerosas tiene derecho a tal ayuda, sin que se especifique ninguna condición con respecto a su nivel de renta. Sin embargo, a raíz de que se haya conocido que varios miembros de la clase política, con rentas altas, y con familia numerosa, eran beneficiarios de tales ayudas, el Gobierno se ha planteado revisar las condiciones para la concesión, exigiendo un determinado umbral de renta. Ello pone sobre la

Esta heterogeneidad en la interpretación y aplicación de la ley de no discriminación de la UE, aunque deseable porque se ajusta y adapta a cada caso, por otro lado, plantea un problema para incorporar consideraciones de equidad y no-discriminación en los sistemas automatizados. Si bien existen numerosas métricas estadísticas en la literatura técnica, ninguna puede capturar de manera confiable una conceptualización europea de la discriminación que es, por definición, contextual. La flexibilidad interpretativa judicial no es un "error" de la ley de no discriminación de la UE; más bien, es a la vez intencional y ventajosa. No obstante, en este entorno regulatorio, la agilidad contextual puede convertirse rápidamente en una subjetividad desbocada, inclinando drásticamente la balanza de poder a favor de posibles discriminadores libres para diseñar sistemas y pruebas de acuerdo con las métricas de *fairness* que les sean más favorables. La igualdad contextual no se presta a la automatización.

En comparación con la toma de decisiones humana, los algoritmos no son tan intuitivos; operan a velocidades, escalas y niveles de complejidad que desafían la comprensión humana, agrupan y actúan sobre clases de personas que no tienen por qué parecerse a grupos históricamente protegidos, y lo hacen sin que las potenciales víctimas sean conscientes del alcance y los efectos de las decisiones automatizadas. La visibilidad de tal discriminación puede pasar fácilmente desapercibida. Como resultado, es posible que las personas a quienes se les haya aplicado un proceso predictivo y/o decisonal automatizado, nunca se den cuenta de

mesa la cuestión de la "interseccionalidad" la cual, por defecto, aquí no se ha dado (familia numerosa + máximo de renta). Y lleva también a poner de relieve que el concepto de "persona vulnerable" no está precisado jurídicamente. En este caso, además, se parte de que el Programa Bosco, que es el que asigna en primer lugar la concesión de ayudas del Bono Social Eléctrico -requisito determinante y previo para ser acreedor del Bono Social Térmico- está puesto en cuestión por su aparente discriminación. Si a ello se suma la falta de toma en consideración por parte de la normativa, de los diversas características y atributos que inciden en la interseccionalidad en el ámbito *offline*, con mayor razón se proyecta potenciado en el ámbito algorítmico.

que han estado en desventaja y, por lo tanto, no tengan un punto de partida para poder presentar una reclamación contra tal discriminación.

La discriminación que deriva de un proceso predictivo o de un proceso decisional de un algoritmo, es decir, una discriminación automatizada, es más abstracta y poco intuitiva, mucho más sutil, intangible y difícil de detectar que cuando es un ser humano quien discrimina. Los seres humanos discriminan debido a actitudes negativas (por ejemplo, estereotipos, prejuicios) y sesgos muchas veces no intencionales (por ejemplo, prácticas organizacionales o estereotipos internalizados) que pueden actuar como una señal para las víctimas, de manera que las permite detectar más fácilmente que se ha producido discriminación. La demanda sólo puede plantearse si una víctima realmente se siente en desventaja. Por ejemplo, un solicitante puede no saber cuándo o por qué se filtró su *curriculum vitae* al solicitar un trabajo. Es posible que el consumidor nunca sepa que no se le muestran ciertos anuncios o que están recibiendo ofertas de productos comparativamente desfavorables con respecto a los que le llegan a otros individuos.

Estos mecanismos de señalización, de advertencia de la discriminación, no existen en los sistemas algorítmicos. Por tanto, habría que diseñar e implementar procedimientos de evaluación consistentes en definir un estándar común para la evidencia estadística, que permitiera detectar y evaluar *prima facie* la discriminación automatizada, y que apoyaran las respectivas funciones, desde los diseñadores de los programas a los controladores, así como a los ulteriores demandantes y a los propios jueces (Wachter et al., 2021). Esta medida estadística de referencia estándar acogería las numerosas nociones que se utilizan para definir la *fairness* y la lucha contra la discriminación, que se nutren de varias propuestas (ceguera voluntaria, paridad demográfica o estadística, igualdad predictiva (Dhasarathy et al.,

2021); *individual fairness, group fairness, fairness through unawareness, equal outcome, equal opportunity, equal impact*, (Barocas, Hardt y Narayanan, 2018); o bien, tratamiento desigual, impacto dispar (Rubin; Verma, 2018), etc.⁸

En definitiva, para prevenir casos que involucren una potencial discriminación causada por sistemas automatizados, habría que encontrar métricas de equidad adecuadas ajustadas a la legalidad y, sin embargo, que sean lo suficientemente estáticas como para ser codificadas. En general, los tribunales nacionales y europeos, así como los estudiosos del derecho, son reacios a depender en gran medida de la evidencia estadística para probar las denuncias de discriminación. Las estadísticas a menudo no están disponibles, son obsoletas o no se consideran confiables.

Por último, a esta dificultad de encontrar un estándar para automatizar la *fairness*, hay que añadir que la discriminación aditiva e interseccional complica aún más estas decisiones automatizadas (por ejemplo, una discriminación, como pueda ser la de género en el lugar de trabajo o prácticas de contratación racistas). ¿Cómo mediría un algoritmo atributos que, superpuestos, da lugar cada una de ellas a un mayor grado de discriminación? Un algoritmo que se empleara para evaluar currícula y que tuviera sesgos contra el género femenino, contra la raza (negra o asiática, por ejemplo), contra la edad (rechazando a solicitantes de más de 40 años),

⁸ Wachter et al. proponen que se aporten una definiciones clave, y que la posición probatoria y los umbrales comparativos se definan en función de las circunstancias del caso, incluyendo (1) quién es un grupo desfavorecido y de comparación apropiado;(2) que es lo que constituye una desventaja particular (por ejemplo, oportunidad financiera perdida), (3) la relevancia del daño discriminatorio (por ejemplo, debe afectar a un “lejano mayor número” de personas, no debe ser fortuito, reglas menos estrictas si son constantes disparidad durante un largo período), (4) si un daño concreto necesita ser probado o una amenaza potencial es suficiente, y (5) qué tipos de evidencia son admisibles en los tribunales (por ejemplo, estadísticas o evaluaciones de sentido común).

contra las personas con necesidades especiales (un cierto grado de minusvalía), en su decisión automatizada de “rechazar” la solicitud de la candidata, no se podría conocer si obedecía a un compendio de todos los atributos mencionados, o con solo uno de ellos ya habría sido suficiente para rechazarlo. En cualquier caso, la dimensión interseccional, amplifica las posibilidades de discriminación automatizada y de la *black box*, ya que no se sabría la razón (razones) concretas.

Ante una decisión automatizada discriminatoria, que afectara a la dimensión de interseccionalidad, el poder judicial, para valorar las pruebas, podrían incentivar lo que comúnmente se conoce como “dividir y conquistar el enfoque”. Es decir, se podría defender el comportamiento discriminatorio de un sistema al dividir al grupo desfavorecido en subgrupos y mostrando que cada subgrupo es tan pequeño que ninguno de ellos pasa el umbral legalmente requerido de desventaja significativa y desproporcionada (es decir, dominancia negativa). Esto sería especialmente perjudicial para casos que involucren discriminación interseccional o grupos minoritarios (Wachter et al. 2021).

CONCLUSIONES

Los procedimientos automatizados a través del aprendizaje automático de una IA deben ajustarse a la *fairness*, al igual que se hace en el ámbito *offline*. Tal desiderátum no es de fácil consecución ya que los sistemas automatizados, especialmente en cuanto a cuestiones de justicia y equidad, encuentran serias dificultades técnicas. En este trabajo he señalado dos de estas problemáticas.

En primer lugar, cuando un algoritmo trabaja con un conjunto de identidades o atributos sensibles (sexo, raza, clase) que influyen sobre el acceso que se pueda tener a derechos y oportunidades, hay que ser especialmente vigilante para que la actividad

predictiva y/o el proceso decisional no vulnere la igualdad y el principio de no-discriminación. El análisis interseccional tiene como objetivo revelar las variadas identidades, exponer los diferentes tipos de discriminación y de desventaja que se dan como consecuencia de una combinación y cruce de identidades. Busca abordar las formas en las que el racismo, el patriarcado, la opresión de clase, y otros sistemas de discriminación crean desigualdades.

En segundo lugar, resulta difícil encontrar un estándar para automatizar la *fairness*, unas métricas de equidad adecuadas que respeten la legalidad y que, sin embargo, sean lo suficientemente estáticas como para ser codificadas y, de esa manera, aplicadas por el algoritmo.

Ahora bien, lo que técnicamente parece inviable ahora, habrá voces que señalen que, posiblemente, en un futuro próximo, como resultado de una investigación interdisciplinar (técnicos, juristas, éticos) puede que sea posible y contemos con una IA que “se aproxime” al criterio de *fairness* de un ser humano. Tengo mis serias reservas sobre tal “acercamiento” ya que el sentido de la justicia, como el de la equidad, precisan de elementos tales como el razonamiento, uso del lenguaje, capacidad crítica, corrección de errores, explicaciones (y no mera explicabilidad), finura no sólo de técnica legislativa sino de intuición, de valores, de justicia, de interpretación y de contexto que, sólo la razón humana está en condiciones de elaborar.

BIBLIOGRAFÍA

- Aristóteles, *Ética a Nicómaco*, Libro V, cap. 3.
Barrere Unzueta, Ma. Ángeles (2010). “La interseccionalidad como desafío al *mainstreaming* de género en las políticas públicas”, *R.V.A.P.* no. 87-88, pp. 225-252.
Barocas, Solon; Selbst, Andrew (2016). “Big Data’s Disparate Impact”, *California Law Review*, v. 104, n. 3, pp. 671-732.

Barocas, Solon; Hardt, Moritz; Narayanan, Arvind (2019). *Fairness and Machine Learning: Limitations and Opportunities*. fairmlbook.org. Disponible en: <https://fairmlbook.org/>. Acceso 12/11/2022

Belloso Martín, Nuria (2023). "Sobre *fairness* y *machine learning*: el algoritmo ¿puede (y debe) ser justo?, *Anales de la Cátedra Francisco Suárez*, Universidad de Granada, nº 57, 2023, pp. 7-38. ISSN: 0008-7750. DOI: <https://doi.org/10.30827/acfs.v57i.25250>
<https://revistaseug.ugr.es/index.php/acfs/article/view/25250>

Beltramelli Gula, Noelia; Camila Ferro, María Goñi Mazzitelli, Lorena Etcheverry & Martín Rocamora (2022). "Un concepto viajero: comprensiones acerca de la interseccionalidad en estudios de sesgo en la ciencia de datos", *NORUS*, vol. 10, nº 18, pp. 152-18. <file:///C:/Users/Propietario/Downloads/4847-Texto%20do%20artigo-8672-2-10-20230119-1.pdf>

Crenshaw, Kimberle (1989). "Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics", *Review University of Chicago Legal Forum*.

Crenshaw, Kimberle (1991). "Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color", *Stanford Law Review*, nº43(6), pp. 1241-1299. <https://doi.org/10.2307/1229039>

Chouldechova, Alexandra (2017). Fair Prediction with Disparate Impact: A study of Bias in recidivism prediction instruments, *Big Data*, v. 5, n. 2, pp. 153-163.

Castilla, Karlos (2022). *Cuatro ángulos de análisis de la igualdad y la no discriminación en la inteligencia artificial*. Institut de Drets Humans de Catalunya.

Castilla, Karlos (2022). "Ideas para el análisis de los derechos humanos desde una perspectiva interseccional", *Papeles el tiempo de los derechos*, nº2, Instituto de Derechos Humanos Universidad Carlos III de Madrid, pp. 1-14.

Dhasarathy, A., Jain, S. & Khan, N. (2020). *When governments turn to AI: Algorithms, trade-offs, and trust*. McKinsey & Company.

Ghosh, Avijit; Genuit, Lea; Reagan, Marry (2021). "Characterizing Intersectional Group Fairness with Worst-Case Comparisons", 2º Workshop on diversity in Artificial Intelligence (aidbei). *Proceedings of Machine Learning Research*, n.142, pp. 22-34.

Stewart, Matthew (2020). Cómo lograr la equidad en los algoritmos. <https://www.codetd.com/es/article/12010244>

Verma, Sahil; Rubin, Julia (2018). "Fairness Definitions Explained", 2018 ACM/IEEE *International Workshop on Software Fairness*.
Young, Iris Marion (2000). *La justicia y la política de la diferencia*, Trad. de Silvina Álvarez, Madrid, Ediciones Cátedra.

JURISPRUDENCIA

Tribunal Europeo de Derechos Humanos. (2012). Caso B.S. vs. España. (No. 47159/08), España: Sentencia de 24 de julio de 2012.
Corte Interamericana de Derechos Humanos. (2015). Caso González Llu y otros vs. Ecuador. Excepciones Preliminares, Fondo, Reparaciones y Costas. Sentencia de 1 de septiembre de 2015. Serie C No. 298.

LA INTELIGENCIA ARTIFICIAL: EL NUEVO AVANCE TECNOLÓGICO CON REPERCUSIÓN EN LOS SISTEMAS JURÍDICOS

Susana Álvarez González
Universidade de Vigo

UN INTENTO DE APROXIMACIÓN AL CONCEPTO DE IA

Dar respuesta a la pregunta qué es la inteligencia artificial (IA) resulta una tarea compleja. McCarthy, la definió en 1955 como el “hacer que una máquina se comporte de formas que serían llamadas inteligentes si un ser humano hiciera eso”, aclarando años más tarde que la IA no se limita a métodos que sean observables biológicamente”. La Real Academia de la Lengua Española define la inteligencia artificial como la “disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico”².

No es éste un concepto pacífico, tanto por la dificultad de determinar lo que se considera conducta inteligente, como por su pretendida comparación con la inteligencia humana, al “modo de

¹ En 2004 la define en los siguientes términos: “It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable. McCARTHY, J.: “What is Artificial Intelligence? *Stanford University*, 24 de noviembre de 2004.

² <https://dle.rae.es/inteligencia>. Consultado el 31 de marzo de 2023.

operar de la mente humana”³. Así, autores como Kaplan, señalan que “hay pocas razones para creer que la inteligencia maquinaal tenga demasiada relación con la inteligencia humana”⁴, sino que se refiere a la creación de programas o máquinas capaces de conductas que podría realizar un ser humano. De hecho, como indica Bostrom, la inteligencia de las máquinas no tiene necesariamente que parecerse a una mente humana⁵. De ahí, que algunos duden sobre si se pueden utilizar las capacidades humanas para valorar o medir la IA⁶.

Como señala Pinto Fontanillo en la actualidad los modelos de IA se mueven entre dos coordenadas, enseñar a ejecutar y a pensar⁷. De ambas, la segunda pretensión señalada resulta sumamente ambiciosa. En este contexto, cabe advertir que no existe un único modelo de IA. La Comisión Europea la definió en sus textos sobre la materia como la combinación de tecnologías que agrupa datos, algoritmos y capacidad informática. En su Comunicación de 25 de abril de 2018, sobre *Inteligencia Artificial para Europa* señalaba que el término se aplica a los sistemas que manifiestan un comportamiento inteligente, pues son “capaces de analizar su entorno y pasar a la acción -con cierto grado de autonomía- con el fin de alcanzar objetivos específicos”, destacando que los sistemas basados en la IA pueden consistir simplemente en un programa

³ PINTO FONTANILLO, J.A.; *El Derecho ante los retos de la inteligencia artificial. Marco ético y jurídico*, Edisofer, 2020, p. 17.

⁴ KAPLAN, J.; *Inteligencia artificial. Lo que todo el mundo debe saber*, TEELL, 2017, p. 1.

⁵ BOSTROM, N.; *Superinteligencia, camino, peligros, estrategias Teell*, 3ª edición, 2016, pp. 28-29.

⁶ “Para entender mejor cómo la ambiciosa conexión entre la máquina y la inteligencia humana enturbia y da color a nuestra comprensión de esta importante tecnología, imaginemos la confusión y polémica que el vuelo a motor podría haber sufrido si los aviones se hubiesen descrito como pájaros artificiales”. KAPLAN, J.; *Inteligencia artificial. Lo que todo el mundo debe saber*, ob.cit., p. 4 y 19.

⁷ PINTO FONTANILLO, J.A.; *El Derecho ante los retos de la inteligencia artificial. Marco ético y jurídico*, ob.cit., p. 24.

informático, pero también ésta puede estar incorporada en dispositivos de hardware.

El alcance y el ámbito de aplicación de la IA puede ser muy diverso por lo que se suelen diferenciar tres categorías: las inteligencias artificiales fuertes, generales y débiles. No obstante, el tipo de IA que ha disparado la aplicación práctica de esta disciplina es la que se conoce como IA-débil (AIweak) y que, en contraste con la IA fuerte y general, se caracteriza por desarrollar soluciones capaces de resolver un problema concreto y acotado. Sobre el desarrollo de la IA general y fuerte existen grandes expectativas que resurgen tras ciertos periodos de hibernación de la disciplina por falta de avances y que dan lugar a un cierto debate social, político y jurídico sobre su regulación. Así, dicha clasificación no está exenta de polémica generando literatura científica dividida entre partidarios y detractores de la misma. Sin embargo, sí parecen existir unos parámetros que, según Bostrom caracterizarán a la IA fuerte: capacidad para aprender; habilidad para manejarse frente a la “incertidumbre e información probable” y algunas capacidades para extraer conceptos adquiridos “dentro de representaciones combinadas usadas en el razonamiento lógico e intuitivo”⁸. Estas características distan de ser pacíficas y, al respecto, señala Lledó Yagüe, que “no es posible que la máquina inteligente pueda alcanzar o implementar ningún proceso cognitivo, mucho menos la mente y sus funciones”⁹.

De todas estas habilidades, las que más inquietud social generan son las vinculadas a la IA cuya finalidad es conseguir casi cualquier

⁸ BOSTROM, N.; *Superinteligencia, camino, peligros, estrategias Teell*, ob.cit., p. 23.

⁹ LLEDÓ YAGÜE, F.; “Realidades y disrupciones en torno al robot inteligente” (el futuro del hombre quebradizo), en LLEDÓ YAGÜE, F., BENÍTEZ ORTÚZAR, I. y MONJE BALMASEDA, O.; *La robótica y la inteligencia artificial en la nueva era de la revolución industrial 4.0 (Los desafíos jurídicos, éticos y tecnológicos de los robots inteligentes*, Dykinson, Madrid, 2021, p. 33.

objetivo, incluido el de aprender¹⁰ y, en particular, el aprendizaje automático. En lo que algunos autores han denominado “aprendizaje maquinal” éste procede de la práctica, el entrenamiento y la experiencia. La vertiente nueva es la utilización y análisis de información a gran escala y de las nuevas técnicas de computación que, en cierto modo, pretenden imitar algunos aspectos del cerebro humano¹¹, con la idea de lograr en algún momento la “singularidad, esto es, que las máquinas puedan ser capaces de rehacerse y mejorarse a sí mismas¹². Si bien es cierto que la posibilidad de crear una máquina “ultrainteligente” en los términos expuestos por Bostrom pudiera implicar una serie de riesgos extremos incluso para la supervivencia de los seres humanos, resulta también importante señalar que quizás estos puedan ser previstos y planificados antes del avance técnico si se concluye que pudiera producirse. La cuestión no deja de ser un asunto especialmente complicado teniendo en cuenta, entre otros, el despliegue y uso del denominado Big Data y sus aplicaciones¹³.

Las últimas modificaciones, en relación con 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, parecen optar por incluir una reestructuración de la definición de los sistemas de inteligencia artificial con el objetivo de distinguirlos de otros más sencillos, centrándose en los “sistemas desarrollados mediante

¹⁰ Es necesario tener en cuenta que la percepción humana sobre la dificultad en la realización de objetivos es engañosa en su aplicación a la IA.

¹¹ KAPLAN, J.; *Inteligencia artificial. Lo que todo el mundo debe saber*, ob.cit., pp. 33 y ss.

¹² *Ibidem*, p. 148.

¹³ Vid. GARRIGA DOMINGUEZ, A.; *Nuevos retos para la protección de datos personales: en la era del big data y de la computación ubicua*, Dykinson, Madrid, 2015.

estrategias de aprendizaje automático y estrategias basadas en la lógica y el conocimiento”¹⁴.

NUEVO AVANCE TÉCNICO, ¿NUEVO DERECHO?

Al margen de discusiones doctrinales en torno a su definición, no cabe duda de que el desarrollo de la IA está marcando en los últimos tiempos la investigación tecnológica, de forma muy similar a lo que sucedió años antes con el Proyecto Genoma Humano, si bien con parámetros y consecuencias distintas¹⁵. Al igual que sucedía entonces, este avance técnico produce una paradoja en relación con los derechos, pues se trata de proyectos fruto de la libertad, investigación y producción científica que abren el camino a múltiples beneficios pero que pueden constituir una amenaza para la persona y sus derechos. Cabe esperar, por tanto, que genere, como mínimo, “la necesidad de nuevas protecciones y de nuevos derechos”¹⁶, pues suele existir correlación entre “avance técnico-cambio jurídico” cuando éste tiene una proyección social amplia.

La consecuencia jurídica del avance técnico puede consistir, como señala Díez-Picazo, en la aparición de una nueva legislación o en la reacción del ordenamiento jurídico en sus vías de aplicación e interpretación, siendo frecuente que la primera reacción sea la de “tratar de ajustar o englobar el conjunto de los nuevos hechos que

¹⁴ <https://www.consilium.europa.eu/es/press/press-releases/2022/12/06/artificial-intelligence-act-council-calls-for-promoting-safe-ai-that-respects-fundamental-rights/>

¹⁵ Algunas de las ideas expuestas en estos apartados han sido previamente incluidas en: ÁLVAREZ GONZÁLEZ, S.; “La necesaria protección de los derechos fundamentales como punto de partida en las propuestas de regulación de la inteligencia artificial”, en *Interfaces dos direitos humanos no século XXI*, Editora Metrics, Brasil, 2022.

¹⁶ PECES-BARBA MARTÍNEZ, G.; “La libertad del hombre y el Genoma”, en *El Derecho ante el Proyecto Genoma*, Vol. I, Fundación BBV con la colaboración de Universidad de Deusto-Diputación Foral de Biskaia, Bilbao, 1994, P. 202.

el nivel técnico determina en los moldes antiguos, preconizando de este modo el mantenimiento de la misma solución anteriormente establecida, por medio de lo que puede ser una forzada interpretación extensiva de los textos”¹⁷. De esta forma, como el Derecho surge de la realidad, es necesario que intervenga con reglas de juego que se impongan en la sociedad en transformación.

Como indica Garriga Domínguez, respecto a la influencia de las TIC, “desde el último tercio del siglo pasado se produce un cambio de paradigma que afecta a todos los aspectos de la sociedad postindustrial” en la cual la información se convierte en fuente fundamental de productividad y también de poder¹⁸. Al auge de la sociedad de la información, hay que adicionar, el avance de otras ciencias y tecnologías. Vinculados a estos retos, surgen nuevos campos de la ética y también del Derecho. En particular, en relación con la IA, hace ya algunos años que la literatura ha señalado, como uno de los grandes retos, la inclusión de parámetros éticos de las máquinas y su grado de cumplimiento¹⁹.

En este contexto, como señala Pérez Luño, el Derecho no existe ni es imaginable fuera del tiempo, de tal forma que los conceptos y categorías del actual surgieron en la Modernidad y en la etapa de codificación con un fin: dar respuestas jurídicas a las exigencias y a los problemas políticos de aquella coyuntura histórica. La realidad ha variado, entre otras cuestiones por la revolución científica y técnica, “por eso, el método de análisis de esa realidad tiene que adecuarse a las nuevas exigencias de su objeto”. El Derecho

¹⁷ DÍEZ-PICAZO, L.; Derecho y masificación social. Tecnología y Derecho privado (Dos esbozos). Cuadernos Civitas, Madrid, 1987, p. 90.

¹⁸ GARRIGA DOMÍNGUEZ, A.; “El impacto de la TIC en los derechos humanos”, en *Fundamentos éticos y jurídicos de las TIC*, Aranzadi, Pamplona, 2012.

¹⁹ De forma paralela surge el debate en el ámbito filosófico sobre la mejora de los aspectos morales de la propia naturaleza humana, Vid. PERSSON, I. and SAVULESCU, J.; *¿Preparados para el futuro?*, TEELL, 2020.

debería tener en cuenta esta nueva realidad si no quiere volverse “inútil o, en el peor de los casos, deformadora de su significado”²⁰.

Al igual que de forma previa sucedió con los avances de la Genética y de la Biotecnología²¹, el avance de la IA debe constituir un nuevo marco de estudio para el Derecho y especialmente para el estudio de los derechos humanos. Partiendo de esa base, y al margen de los posibles ilícitos penales, los derechos fundamentales con más riesgo de afección por el uso de esta tecnología son el derecho a la igualdad y no discriminación y el derecho a la privacidad. Al respecto, cabe recordar que este último, también denominado derecho a la autodeterminación informativa o protección de datos, constituye una nueva modalidad de libertad personal, tendente a proteger jurídicamente la “identidad personal”²². Es la respuesta del Estado de Derecho frente a las nuevas necesidades derivadas de la revolución tecnológica y que garantiza a la persona un poder de control y disposición sobre sus propios datos personales. Este actúa al mismo tiempo como instituto o garantía del pleno disfrute de los restantes derechos fundamentales, como pueden ser la igualdad y la no discriminación²³.

Si bien los derechos señalados son los que pueden verse directamente más afectados por la creación y/o usos de la IA según el estado actual de la técnica, ésta presenta también un riesgo directo o indirecto para otros derechos que pueden verse comprometidos en la deriva de un uso torticero de la tecnología. Así, si los riesgos inmediatos deben abordarse con normas para proteger la privacidad, la igualdad y la no discriminación, resulta

²⁰ PÉREZ LUÑO, A.E.; *Trayectorias contemporáneas de la Filosofía y Teoría del Derecho*, Grupo Nacional de Editores, Sevilla, 2003 p. 4.

²¹ Vid. ÁLVAREZ GONZÁLEZ, S.; *Derechos fundamentales y protección de datos genéticos*, Dykinson, 2007.

²² PÉREZ LUÑO, A.E.; “Vittorio Frosini y los nuevos derechos en la sociedad tecnológica”, *Informática e Diritto*, Le Monnier, Firenze, 1992, p. 107.

²³ Fundamento jurídico séptimo, sentencia del Tribunal Constitucional Español 290/2000, de 30 de noviembre.

también esencial la valoración de los posibles daños para otras libertades. Entre los derechos a los que debe prestarse una especial atención ante un mal uso o creación de IA, cabe señalar la libertad de expresión, la libertad de reunión, la libertad de pensamiento, el derecho a la tutela judicial efectiva y a un juicio justo, el derecho a la educación e incluso el derecho a la vida y a la salud. Muchos de estos derechos ya están siendo afectados por el uso de ciertas aplicaciones con fines diversos.

El Consejo de Europa aprobó el 8 de abril de 2020 una serie de *Recomendaciones sobre el impacto de los algoritmos en los derechos humanos*. Se insta a los Estados miembros a garantizar el derecho a la autodeterminación informativa en el diseño, desarrollo y puesta en marcha de los sistemas algorítmicos, evaluando con detalle los derechos humanos que pueden verse afectados y, en particular, el derecho a la no discriminación. Se insta, de forma específica, al sector privado a garantizar los derechos humanos tomando conciencia de la responsabilidad hacia los valores de una sociedad democrática. En este sentido, cabe recordar, con el profesor Peces-Barba, que, si bien los derechos fundamentales se presentan en su fórmula clásica como mecanismos para preservar la libertad del individuo frente al poder de Estado, para proteger a los hombres “de los maleficios del poder”, la quiebra de la convicción igualitaria ha promovido la extensión de los derechos fundamentales al ámbito privado, de tal forma que ha permitido preguntarse si los derechos no deben ser también preferentes o, en todo caso, protegibles en las mismas condiciones en los dos ámbitos²⁴.

Aunque es imposible predecir cómo va a evolucionar el estado, los conocimientos y aplicaciones vinculadas a la IA en los próximos años, parece indudable que el Derecho deberá enfrentarse a nuevos retos derivados del análisis de grandes cantidades de datos de carácter personal y su conexión, incluso con relación a

²⁴ PECES-BARBA MARTÍNEZ, G., *Ética, poder y Derecho. Reflexiones ante el fin de siglo*, Centro de Estudios Constitucionales, Madrid, 1995, p. 72.

conjuntos de datos que, en sí mismos, no contienen datos personales. Previsiblemente, si se logran grandes avances en la inteligencia artificial general y fuerte, el encaje de los riesgos en las categorías actuales de derechos resulte extremadamente complejo, siendo preciso quizás una extensión de su contenido que dé cobertura a nuevas situaciones o tal vez la creación de nuevos derechos²⁵. Los poderes públicos deben estar preparados para abordar con “el nivel superior de maestría necesario” la competencia tecnológica²⁶. En este sentido, como señala Llano Alonso, “debemos y podemos salvar nuestra circunstancia con las tecnologías, pero a la vez debemos librarnos de su fuerza de atracción centrípeta para evitar ser absorbidos por el agujero negro del posthumanismo”²⁷.

LA CONFIANZA SOCIAL: LA NECESIDAD DE COMPRENSIÓN, ACEPTACIÓN Y EL PRINCIPIO DE TRANSPARENCIA COMO REQUISITO ESENCIAL EN LA REGULACIÓN DE LA IA.

Como señala Salazar, la expansión exponencial de la tecnología pasa, en última instancia, por la aceptación social y comprensión de la misma, ante un cambio tecnológico tan rápido que resulta difícil de asumir y que puede provocar el denominado “shock tecnológico”²⁸.

Tras ciertos escándalos por todos conocidos derivados de un uso masivo ilícito de información de carácter personal con afección

²⁵ Sobre los nuevos derechos y tecnologías, Vid. DE ASÍS ROIG, R., *Derechos y tecnologías*, Dykinson, Madrid, 2022, pp. 43-51.

²⁶ BOSTROM, N.; *Superinteligencia. Caminos, peligros, estrategia*, ob.cit., p. 11.

²⁷ LLANO ALONSO, F.H.; *Homo Excelsior. Los límites ético-jurídicos del transhumanismo*, Tirant lo Blanch, Valencia, 2018, p. 192.

²⁸ “mezcla de miedo a lo desconocido y una vaga conciencia de sus ventajas e inconvenientes”. SALAZAR, I.; *La revolución de los robots. Cómo la inteligencia artificial y la robótica afectan a nuestro futuro*, Ediciones TREA, Madrid, 2019, p. 51.

directa a los derechos fundamentales, parece que, en gran medida, la confianza en sistemas vinculados al uso de la IA, dentro del marco del Estado de Derecho, vendrá dada y quedará condicionada por la forma en la que se realice su regulación y por las garantías que se implementen para la protección de estos derechos. La Comisión europea ha señalado como uno de los objetivos centrales en el abordaje de la materia la necesidad de generar un “ecosistema de confianza” con un enfoque antropocéntrico centrado en el ser humano. Dicho reto pasa por la necesidad previa de una base mínima de alfabetización digital de la ciudadanía, pues, malamente, podrá valorar un individuo las bondades o maldades de la implantación tecnológica o de la toma de decisiones basada en la misma si no le resulta posible su comprensión. Desde algunos sectores se ha apelado a una cierta “responsabilidad de la ciudadanía” que no se estaría ejerciendo en relación con los avances tecnológicos. No obstante, la misma no parece exigible en un contexto actual en el que la educación y la falta de comprensión de esta materia no permiten este control o responsabilidad en su utilización. Al respecto, la Unión Europea, dentro del marco de su *Estrategia sobre IA*, aprobada en abril de 2018, ya destacó la necesidad de concienciación en torno a la misma, en todos los niveles de la educación, con la finalidad de crear habilidades sobre la toma de decisiones. Evidentemente, dicha tarea resultará de enorme complejidad y de difícil extensión a ciertos sectores concretos de la población.

Los recelos sobre el uso tecnológico mal intencionado o sobre los posibles efectos perversos parecen forzar un debate técnico o pretécnico. En este contexto, la Comisión Europea, de forma previa a entrar en la regulación, ha señalado, siete requisitos a tener en cuenta en relación con la IA: acción y supervisión humanas; solidez técnica y seguridad; gestión de la privacidad y protección de datos; transparencia; diversidad, no discriminación y equidad; bienestar social y medioambiental y rendición de cuentas. Asume las *Directrices éticas para una IA fiable* elaboradas por el grupo independiente de expertos de alto nivel sobre inteligencia artificial, creado en junio de 2018.

Entre ellos, la transparencia se presenta como un requisito esencial no solo respecto a los datos de entrenamiento o a la conservación de los datos sino también para la creación de esa confianza social e individual y, en su caso, para reparar el daño causado. Según el *Libro Blanco de IA*, este principio de transparencia conlleva necesariamente: la información respecto a las capacidades y limitaciones de la IA e información sobre la si el afectado se está relacionado con un ser humano o con un sistema de IA.

Este principio de transparencia está ya presente en la normativa de protección de datos en el ámbito de la Unión Europea, considerándolo como una medida de privacidad por defecto para que los ciudadanos puedan supervisar el tratamiento de su información de carácter personal. La normativa vigente en materia de protección de datos lo señala como una medida imprescindible para garantizar un tratamiento leal de los datos personales y del derecho a recibir una información completa, clara y sencilla relativa a todos los aspectos relevantes de un tratamiento de datos personales y a sus posibles consecuencias²⁹.

No obstante, dicho principio ha de complementarse, en el marco de la IA y en aras a generar la confianza social en torno a la misma, con su extensión en el diseño y en la puesta en marcha de los dispositivos que usen este tipo de tecnología. Así, “este derecho cobra una nueva y mayor relevancia en el mundo de los datos masivos, de la computación ubicua y del Internet de las cosas”³⁰. Esta exigencia implicará que la aplicación de dicho principio no puede reducirse a un instante puntual, sino que afectará a todos y cada uno de los momentos de creación y puesta en marcha de los

²⁹ GARRIGA DOMÍNGUEZ, A.; “La elaboración de perfiles y su impacto en los derechos fundamentales. Una primera aproximación a su regulación en el Reglamento de Protección de datos de la Unión Europea”, *Derechos y Libertades*, núm.38, época II, 2018, p. 134.

³⁰ *Ibidem*, p. 136.

dispositivos, así como a todos los elementos y participantes que intervienen en la solución tecnológica.

Este principio debe complementarse necesariamente con la supervisión humana, con el objetivo de no socavar la autonomía humana y protegerla de los efectos adversos. Esta supervisión planteará especial problema en aquellos sistemas de IA que evolucionen y aprendan de la experiencia en los que necesariamente se tendrán que adoptar mecanismos para evitar la pérdida de control de la IA por el ser humano. En el marco de las *Directrices Éticas para una IA fiable* se apuesta por una IA centrada en las personas, en el servicio a la humanidad y en el bien común y apoyada en tres componentes esenciales: lícita, que respete el marco normativo existente, prestando especial atención a los textos de derechos humanos, ética y robusta desde el punto de vista técnico y social para que los sistemas funcionen de manera segura y fiable y determinación, en su caso, los posibles daños involuntarios. Dentro de los principios éticos de la IA, basados en los derechos humanos, el citado grupo de expertos destaca los siguientes: respeto a la autonomía, prevención del daño, equidad y explicabilidad.

El Parlamento Europeo, el Consejo y la Comisión se han pronunciado sobre los parámetros analizados en este epígrafe en la *Declaración Europea sobre los Derechos y Principios Digitales* para la Unión Europea, de 26 de enero de 2022³¹. Si bien el texto tiene carácter declarativo, aborda, entre otras cuestiones, en el capítulo III, las “interacciones con algoritmos y sistemas de inteligencia artificial”, en la que se compromete políticamente, en aras a garantizar la libertad de elección, a la promoción y aplicación de dicho principio. La finalidad última, en este contexto, debería ser evitar que la IA y los algoritmos decidan “sobre fines y objetivos futuros del desarrollo de la vida, al margen de las elecciones de las

³¹<https://digital-strategy.ec.europa.eu/en/library/declaration-european-digital-rights-and-principles#Declaration>

personas concernidas por esos designios” y en ocasiones con sustitución plena de las decisiones humanas por la IA³².

ALGUNOS APUNTES SOBRE UN DEBATE ABIERTO

En esta temática son habituales las denominados pseudocontroversias³³ que se recrean en posturas antagónicas sobre las bondades o maldades de la IA y no en el estudio de las competencias y objetivos que puede alcanzar. Esto no significa que los sistemas jurídicos puedan permanecer ajenos a este avance complejo, con independencia de que finalmente se produzca o no, sino que el debate ha de producirse en el marco de los indicadores señalados. Así, cuando hablamos de IA, es necesario centrar el debate en sus justos términos, entendiendo esta tecnología como una combinación de información y computación, por lo que será necesario movernos en esos parámetros y en el control de los mismos.

Situar el debate en estos términos, no significa que su regulación haya de abordarse una vez que se produce el avance sino más bien al contrario pues, desde el punto de vista jurídico, estos puntos deben ser predefinidos previamente y analizados de conformidad los parámetros del ordenamiento. Esta postura plantea un problema inicial: la necesidad de planificación previa.

En este sentido, “el riesgo real de la IA no es la maldad sino la competencia”³⁴, de tal forma que, cuando se analiza el tema, “debemos ser conscientes de los peligros, identificarlos, emplear

³² PÉREZ LUÑO, A.E.; “El posthumanismo no es humanismo”, *Derechos y Libertades*, núm. 44, enero, 2021, p. 30.

³³ Sobre esta cuestión, Vid. TEGMARK, M.; *Vida 3.0. Que significa ser humano en la era de la inteligencia artificial*, Penguin Random House Grupo Editorial, 2018, pp. 59 y ss.

³⁴ HAWKING, S.; *Breves respuestas a las grandes preguntas*, ob.cit. p. 233.

la mejor práctica y gestión posibles y prepararnos con mucha antelación para sus consecuencias”³⁵.

Si no resulta posible este control, quizás debería ponerse en marcha la elaboración de un Derecho preventivo. Sin embargo, determinar qué usos y qué avances tecnológicos deberían prohibirse resulta, sin duda, complicado habida cuenta del estado actual de las tecnologías emergentes. No son pocos los que apostarían por prohibir aquellas tecnologías en las que el riesgo no sea razonable o proporcional en relación con los beneficios, pero dicha afirmación nos conduciría a la aplicación de principios con un alto grado de incerteza. Es, en este contexto, en el que los derechos humanos, ya anteriormente señalados, crearían un límite, no solo como punto de partida, sino como punto de llegada.

³⁵ *Ibidem*, p.253.

REGULAÇÃO DA INTELIGÊNCIA ARTIFICIAL NO BRASIL

Licurgo Mourão

Pós-doutorando e Doutor em Direito - Universidade de São Paulo.
Conselheiro substituto do Tribunal de Contas de Minas Gerais.

INTRODUÇÃO

Na atualidade, vivenciamos a transformação de toda a humanidade: a forma como vivemos, trabalhamos e nos relacionamos, por meio da expressiva e acelerada revolução tecnológica, que pode ser compreendida como a Quarta Revolução Industrial¹.

O lançamento, pela OpenAI, do ChatGPT, um *software* que simula conversas humanas com usuários e responde a perguntas, deixou ainda mais evidentes o avanço dos estudos na área e os impactos iminentes e drásticos que novas tecnologias podem ocasionar, além de demonstrar a imprevisibilidade do potencial de transformação tecnológica a longo prazo.

A inteligência artificial, que pode ser conceituada como a capacidade de raciocínio de máquinas que reproduz o funcionamento da mente humana, tem ganhado cada vez mais espaço em diversas áreas da atuação humana: desde tarefas corriqueiras, como a influência de algoritmos empregados nas redes sociais, que sugerem com quem fazer amizade e quais conteúdos consumir nas plataformas de *streaming*, até processos automatizados de tomada de decisão, que já interferem em diversos setores.

¹ SCHWAB, Klaus. **A quarta revolução industrial**. Tradução de Daniel Moreira Miranda. São Paulo: Edipro, 2016.

Em outra vertente, o progresso da inteligência artificial pode trazer consigo riscos para os direitos humanos sob o ponto de vista de suas implicações sociais, éticas e jurídicas, tanto da perspectiva individual quanto da coletiva.² Os sistemas de inteligência artificial não podem se sobrepor aos direitos dos cidadãos, sendo necessário que as novas tecnologias sejam desenvolvidas para atuar de forma ética, transparente, sem reproduzir e reforçar as desigualdades e preconceitos presentes na sociedade.

Ao tratar dos desafios da Quarta Revolução Industrial, Klaus Schwab³ enumera os impactos positivos da utilização da inteligência artificial para ações tais como a tomada de decisões racionais orientadas por dados, ganhos no trabalho e em inovação, reorganização de burocracias ultrapassadas e avanços na ciência médica, com a erradicação de doenças. Por outro lado, o autor alerta para problemas relacionados à prestação de contas e responsabilização, perda de trabalho, aumento da desigualdade e novas modalidades de crimes.

Diante das mudanças complexas e velozes acarretadas pelas novas tecnologias, o mencionado autor defende a necessidade de que a Quarta Revolução Industrial seja moldada pela atuação de todos os setores (acadêmico, social, político, nacional e industrial) para garantir que ela seja “empoderadora e centrada no ser humano”, ao invés de “divisionista e desumana”.⁴

Em vista disso, a Organização para Cooperação e Desenvolvimento Econômico (OCDE), ao reconhecer que a inteligência artificial está transformando sociedades, setores econômicos e o futuro do trabalho e que possui potencial de

² GORZONI, Paula. **Inteligência artificial: riscos para direitos humanos e possíveis ações.** Disponível em: <https://itsrio.org/wp-content/uploads/2019/03/Paula-Gorzoni.pdf>. Acesso em: 9 mar. 2023.

³ SCHWAB, Klaus. **A quarta revolução industrial.** Tradução de Daniel Moreira Miranda. São Paulo: Edipro, 2016.

⁴ SCHWAB, Klaus. **Op. cit.**

contribuir para a solução dos principais desafios globais, ressaltou que as transformações acarretadas pelo avanço tecnológico podem afetar desigualmente sociedades e economias, sobretudo no que diz respeito a concorrência, transições no mercado de trabalho, desigualdades e implicações para as democracias e direitos humanos, privacidade e proteção de dados e segurança digital.⁵

Nesse cenário, observa-se a crescente preocupação mundial com a definição dos padrões e limites éticos do emprego da inteligência artificial, buscando regulamentar o uso das novas tecnologias de forma a zelar pelos direitos dos cidadãos.

O grande desafio que se impõe na regulação da inteligência artificial é o de conciliar a integridade na utilização das novas tecnologias, evitando abusos à privacidade e ao exercício de direitos, sem desestimular o avanço das pesquisas e o desenvolvimento de novos sistemas considerados fundamentais, sobretudo no âmbito da Administração Pública, para a contínua melhoria dos serviços prestados à sociedade.

No Brasil, como se verá a seguir, a preocupação com a utilização ética das novas tecnologias ocasionou iniciativas de regulação e projetos de lei, em andamento, que objetivam estabelecer princípios, regras, diretrizes e fundamentos para o desenvolvimento e a aplicação da inteligência artificial.

REGULAÇÃO DA INTELIGÊNCIA ARTIFICIAL E ATUAÇÃO DO ESTADO

Em que pesem todos os benefícios esperados com a revolução tecnológica, não se pode perder de vista que a utilização

⁵ OCDE. **Recommendation of the Council on Artificial Intelligence**. Maio, 2019. Disponível em: <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449>. Acesso em: 9 mar. 2023.

expressiva de inteligência artificial, tanto pelo setor privado quanto pela Administração Pública, pode ocasionar ofensa a direitos individuais e coletivos, motivo pelo qual se mostra necessária a intervenção do Estado por meio da sua atividade regulatória.

Isso porque o ideal de mercado – cuja “mão invisível”, defendida por Adam Smith, acarretaria a melhoria do bem comum por meio da busca dos interesses individuais – mostrou-se insuficiente, diante da ocorrência de falhas de mercado. Assim, em situações de concorrência imperfeita, cabe “ao Estado regular variáveis centrais da atividade econômica, como padrões de qualidade, níveis adequados de informação, condições de entrada e saída do mercado, até limites de preço”.⁶

No caso específico das tecnologias disruptivas, mostra-se relevante a falha de mercado da assimetria informacional, tendo em vista que a ausência de conhecimentos específicos sobre ciências da computação afeta a compreensão dos usuários acerca das tecnologias que impactam sua rotina. Diante da ocorrência de falhas na informação, ensinam Fábio Giambiagi e Ana Cláudia Além, *in verbis*:

[...] a intervenção do Estado justifica-se em razão de o mercado por si só não fornecer dados suficientes para que os consumidores tomem suas decisões racionalmente. A forma de ação do Estado pode ser mediante introdução de uma legislação que induza a uma maior transparência do mercado. [...] a importância da intervenção do Estado não se limita à simples proteção dos consumidores. Tendo em vista que a informação pode ser considerada um bem público – favorecendo todos os agentes do sistema econômico –, o grande objetivo do Estado deve ser

⁶ BINENBOJM, Gustavo. **Poder de polícia, ordenação, regulação**. Belo Horizonte: Fórum, 2020. p. 166.

contribuir para que o fluxo de informações seja o mais eficiente possível.⁷

Observa-se que, em muitos casos, os atingidos pela tecnologia sequer têm ciência de que estão submetidos a ela, a exemplo dos algoritmos que direcionam os conteúdos consumidos nas redes sociais.

Ademais, a tecnologia muitas vezes é criada por determinados grupos de pessoas ou empresas com alguns interesses específicos, podendo até mesmo haver sistemas baseados em inteligência artificial que reproduzem comportamentos discriminatórios presentes na sociedade. “Esses riscos podem ser percebidos quando a inteligência artificial é aplicada na seleção de candidatos em entrevista de emprego ou concessão de crédito por bancos”.⁸

Em geral, modelos computacionais de grande porte e alto custo são desenvolvidos por poucas grandes empresas, localizadas em polos de tecnologia, o que demonstra concentração de poder nas mãos de apenas alguns conglomerados econômicos.

É por isso que o ponto sensível da regulamentação da inteligência artificial diz respeito à necessária transparência e explicabilidade dos sistemas. A transparência deve alcançar o desenvolvimento dos sistemas, os dados utilizados e a decisão tomada com base neles; a explicabilidade corresponde à possibilidade de que o usuário compreenda o modo como a tecnologia chegou ao resultado almejado.

⁷ GIAMBIAGI, Fabio; ALÉM, Ana Claudia. **Finanças públicas: teoria e prática** no Brasil. Rio de Janeiro: Elsevier, 2011. p. 8.

⁸ GORZONI, Paula. **Inteligência artificial: riscos para direitos humanos e possíveis ações**. Disponível em: <https://itsrio.org/wp-content/uploads/2019/03/Paula-Gorzoni.pdf>. Acesso em: 9 mar. 2023.

A melhor forma de se permitir segurança na utilização da inteligência artificial é garantir a compreensão de como determinada decisão foi alcançada. Isso porque as exigências de transparência e explicabilidade das tecnologias reduzem a assimetria informacional e possibilitam a fiscalização do modo como a tecnologia opera e dos dados utilizados na sua atuação. Essa auditoria no funcionamento da tecnologia permite, ainda, a verificação da existência de vieses discriminatórios.

No entanto, as empresas de tecnologia sustentam que as determinações de transparência podem encontrar obstáculos práticos, tendo em vista que, conforme publicado na *MIT Technology Review*, do Instituto de Tecnologia de Massachusetts, o método de redes neurais possui processos tão complexos que nem mesmo seus criadores entendem completamente como chegam às suas conclusões,⁹ além do fato de que o acesso irrestrito aos códigos-fonte pode esbarrar em regras de segredos comerciais, industriais e de propriedade intelectual.

Diante dos desafios observados pelo uso da inteligência artificial, avultam como principais princípios regulatórios que devem nortear a elaboração de leis sobre o tema: a) interesse geral, garantindo inclusão e equidade, uma vez que o poder público não pode visar à satisfação de interesses particulares em detrimento do interesse público; b) direitos humanos, pois devem ser respeitados os direitos fundamentais, tais como as liberdades individuais, a igualdade e a diversidade, combatendo-se práticas discriminatórias realizadas por sistemas de inteligência artificial; c) transparência, de modo a assegurar a compreensão do seu

⁹ HEIKKILÄ, Melissa. **A Quick Guide to the Most Important AI Law You've Never Heard of.** Disponível em:

www.technologyreview.com/2022/05/13/1052223/guide-ai-act-europe/?utm_campaign=site_visitor.unpaid.engagement&utm_medium=tr_social&utm_source=Twitter. Acesso em: 6 mar. 2023.

funcionamento (quanto a algoritmos, dados, provas e validações realizadas, além da lista das áreas que utilizam tais tecnologias).¹⁰

No mesmo sentido, interessante mencionar o estudo realizado pelo Centro de Tecnologia e Sociedade (CTS) da Fundação Getúlio Vargas (FGV Direito Rio), o qual foi encaminhado ao Ministério da Ciência Tecnologia Inovações e Comunicações (MCTIC) no âmbito da consulta sobre a Estratégia Brasileira de Inteligência Artificial. Nesse estudo, em consonância com a Organização para a Cooperação e Desenvolvimento Econômico, recomendou-se que a regulação da inteligência artificial observe os seguintes princípios:

i) os sistemas de IA devem promover o crescimento inclusivo, o desenvolvimento sustentável e o bem-estar;

ii) os sistemas de IA devem ser projetados de maneira a respeitar o Estado de Direito, os direitos humanos, os valores democráticos e a diversidade, liberdade, dignidade, autonomia, privacidade e proteção de dados, não discriminação e igualdade, diversidade, equidade, justiça social e direitos trabalhistas internacionalmente reconhecidos. Para isso, devem incluir salvaguardas apropriadas, como possibilitar a intervenção humana sempre que necessário;

iii) deve haver transparência e explicabilidade sobre os sistemas de IA. Por esse motivo, os atores da IA devem fornecer informações significativas e apropriadas ao contexto, de modo a permitir que os potenciais indivíduos

¹⁰ AGENCIA DE GOBIERNO ELECTRÓNICO Y SOCIEDAD DE LA INFORMACIÓN Y DEL CONOCIMIENTO (AGESIC). **Principios Generales sobre Inteligencia Artificial para un Gobierno Digital**. Disponível em: <https://bit.ly/2Siy4tO>. Acesso em: 24 out. 2022.

afetados tenham acesso aos critérios que serviram de base para a previsão, recomendação ou decisão algorítmica;

iv) os sistemas de IA devem ser robustos, seguros e protegidos durante todo o ciclo de vida. Os riscos potenciais devem ser continuamente avaliados e gerenciados. Para esse fim, os atores da IA devem garantir a rastreabilidade, inclusive em relação aos conjuntos de dados, processos e decisões tomadas durante o ciclo de vida do sistema de IA³, possibilitando a análise dos resultados do sistema e, em caso de solicitação ou investigação, devem oferecer as respostas apropriadas ao contexto e consistentes com o estado da arte; e

v) as organizações e os indivíduos que desenvolvem, implantam ou operam sistemas de IA devem ser responsabilizados pelo seu apropriado funcionamento, de acordo com os princípios acima.¹¹

Imperioso destacar, contudo, que a regulação da inteligência artificial não é tarefa simples, uma vez que, além de buscar garantir a proteção dos direitos do usuário e a segurança na sua utilização, é salutar que se promova e preserve a inovação,¹² de modo a evitar barreiras à pesquisa e ao desenvolvimento de serviços de inteligência artificial. Soma-se a isso a complexidade advinda do cenário de rápida evolução das tecnologias, o que torna a legislação propensa a se tornar obsoleta.

¹¹ IVAR, A. et. al. **Regulação de inteligência artificial no Brasil: policy paper**. 2020. Disponível em: <https://bibliotecadigital.fgv.br/dspace/handle/10438/30078>. Acesso em: 24 out. 2022.

¹² KELLER, Clara Iglesias; BAPTISTA, Patrícia. Por que, quando e como regular as novas tecnologias? Os desafios trazidos pelas inovações disruptivas. **Revista de Direito Administrativo - RDA**, ano 11, n. 273, set./dez. 2016. Disponível em: www.forumconhecimento.com.br/periodico/125/347/3395. Acesso em: 1º mar. 2023.

INICIATIVAS DE REGULAÇÃO DA INTELIGÊNCIA ARTIFICIAL NO BRASIL

No Brasil, algumas legislações tangenciam temas relacionados à inteligência artificial, tais como a Lei n. 12.965/2014, conhecida como Marco Civil da Internet, e a Lei n. 13.709/2018, Lei Geral de Proteção de Dados Pessoais (LGPD), que assegura a explicabilidade de qualquer decisão automatizada que utilize dados pessoais de indivíduos, tendo em vista que as pessoas afetadas por decisões algorítmicas devem poder pleitear uma explicação dos principais critérios e fundamentos utilizados para a decisão automatizada. Contudo, não há legislação federal específica sobre o tema, apenas propostas legislativas em andamento, conforme se verá à frente.

Merece destaque, ainda, a Resolução n. 332/2020, do Conselho Nacional de Justiça,¹³ que estabeleceu diretrizes sobre ética, transparência e governança de inteligência artificial no Poder Judiciário, considerando, sobretudo, a concepção de que a inteligência artificial pode contribuir com a agilidade e coerência do processo de tomada de decisão e que há inúmeras iniciativas nesse sentido nos tribunais.

A Resolução prevê a obrigatoriedade de que o desenvolvimento, a implantação e o uso da inteligência artificial observem os direitos fundamentais e que a utilização de tecnologias deve buscar garantir a segurança jurídica e a igualdade. Para tanto, previu que, antes de ser implementado, o sistema deverá ser homologado, de forma a identificar se preconceitos ou generalizações influenciaram seu desenvolvimento, acarretando tendências

¹³ BRASIL. Conselho Nacional de Justiça. **Resolução n. 332**, de 21 de agosto de 2020. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/3429>. Acesso em: 6 mar. 2023.

discriminatórias no seu funcionamento. Caso não possa ser sanado o viés discriminatório, o modelo não poderá ser utilizado.

Como medida de transparência e controle pelo usuário, os sistemas que auxiliem na elaboração de decisão judicial devem explicar satisfatoriamente a técnica utilizada e os processos que conduziram ao resultado, permitindo a supervisão do magistrado competente.

Embora não se trate propriamente de regulação sobre a matéria, foi instituída a Portaria do Ministério da Ciência, Tecnologia e Inovações n. 4.617, de 6 de abril de 2021, alterada pela Portaria n. 4.979, de 13 de julho de 2021, do mesmo Ministério, a Estratégia Brasileira de Inteligência Artificial (Ebia), que se fundamenta em cinco princípios para uma gestão responsável dos sistemas de inteligência artificial, quais sejam: (i) crescimento inclusivo, desenvolvimento sustentável e bem-estar; (ii) valores centrados no ser humano e na equidade; (iii) transparência e explicabilidade; (iv) robustez, segurança e proteção; e (v) responsabilização ou prestação de contas (*accountability*).

A Estratégia Brasileira de Inteligência Artificial tem como objetivos:

- contribuir para a elaboração de princípios éticos para o desenvolvimento e uso de inteligência artificial responsáveis;
- promover investimentos sustentados em pesquisa e desenvolvimento em inteligência artificial;
- remover barreiras à inovação em inteligência artificial;
- capacitar e formar profissionais para o ecossistema da inteligência artificial;
- estimular a inovação e o desenvolvimento da inteligência artificial brasileira em ambiente internacional; e

- promover ambiente de cooperação entre os entes públicos e privados, a indústria e os centros de pesquisas para o desenvolvimento da inteligência artificial.¹⁴

O referido documento ressalta que, entre as principais discussões envolvendo o uso da inteligência artificial, está a busca por estabelecer princípios gerais e padrões éticos que devem buscar o equilíbrio entre:

(i) a proteção e a salvaguarda de direitos, inclusive aqueles associados à proteção de dados pessoais e à prevenção de discriminação e viés algorítmico;

(ii) a preservação de estruturas adequadas de incentivo ao desenvolvimento de uma tecnologia cujas potencialidades ainda não foram plenamente compreendidas; e

(iii) o estabelecimento de parâmetros legais que confirmam segurança jurídica quanto à responsabilidade dos diferentes atores que participam da cadeia de valor de sistemas autônomos.¹⁵

No final de 2022, a Comissão de Juristas instituída pelo Ato do Presidente do Senado n. 4, de 2022, apresentou minuta de substitutivo aos Projetos de Lei n. 5.051/2019, 21/2020 e 872/2021, que têm como objetivo estabelecer princípios, regras, diretrizes e fundamentos para regular o desenvolvimento e a aplicação da inteligência artificial no Brasil. Nos termos da exposição de motivos,

¹⁴ BRASIL. Ministério da Ciência, Tecnologia e Inovação. **Estratégia Brasileira de Inteligência Artificial (Ebia)**. Disponível em: www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/inteligencia-artificial. Acesso em: 14 mar. 2023.

¹⁵ BRASIL. Ministério da Ciência, Tecnologia e Inovação. **Estratégia Brasileira de Inteligência Artificial (Ebia)**. Disponível em: www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/inteligencia-artificial. Acesso em: 14 mar. 2023.

[...] esse novo marco legal tem um duplo objetivo. De um lado, estabelecer direitos para proteção do elo mais vulnerável em questão, a pessoa natural que já é diariamente impactada por sistemas de inteligência artificial, desde a recomendação de conteúdo e direcionamento de publicidade na Internet até a sua análise de elegibilidade para tomada de crédito e para determinadas políticas públicas. De outro lado, ao dispor de ferramentas de governança e de um arranjo institucional de fiscalização e supervisão, criar condições de previsibilidade acerca da sua interpretação e, em última análise, segurança jurídica para inovação e o desenvolvimento econômico-tecnológico.¹⁶

A Comissão de Juristas realizou, além de reuniões internas, audiências públicas envolvendo mais de 50 (cinquenta) palestrantes, objetivando a participação dos diversos setores da sociedade, como setor público, academia, indústria e terceiro setor de diversas regiões do país. Também foram realizados seminários internacionais com mais de 20 (vinte) palestrantes e, ainda, recebidas 102 (cento e duas) contribuições escritas de entidades representantes da sociedade civil, de órgãos governamentais, da academia, do setor privado, além de contribuições individuais. O relatório final apresentado resultou em um documento de 912 (novecentas e doze) laudas, o que demonstra a complexidade das discussões que envolvem a matéria.

A minuta formulada pela Comissão de Juristas, mais robusta que a proposta inicial que havia sido aprovada na Câmara dos Deputados, possui 45 (quarenta e cinco) artigos que objetivam

¹⁶ BRASIL. Senado Federal. Coordenação de Comissões Especiais, Temporárias e Parlamentares de Inquérito. **Relatório final** da Comissão de Juristas responsável por subsidiar a elaboração de Substitutivo sobre Inteligência Artificial instituída pelo Ato do Presidente do Senado n. 4, de 2022, aprovado em 1º de dezembro de 2022. Disponível em: <https://legis.senado.leg.br/sdleg-getter/documento?dm=9221643&ts=1677521675945&disposition=inline>. Acesso em: 15 mar. 2023.

trazer segurança aos usuários das novas tecnologias, prevendo, em suma, diversos direitos às pessoas afetadas pela inteligência artificial; a possibilidade de responsabilização do desenvolvedor ou operador da tecnologia; a exigência de transparência e medidas de governança, que devem ser observadas ao longo de todo o ciclo de vida do sistema de inteligência artificial, desde a concepção inicial até o encerramento de suas atividades e descontinuação.

O substitutivo adotou modelagem baseada nos riscos que o sistema de inteligência artificial pode ocasionar, prevendo que todo sistema passará por análise preliminar de classificação de seu grau de risco e, a partir daí, são preconizadas regras diferenciadas, de acordo com os potenciais riscos da tecnologia.

Determinou, ainda, a designação, pelo Poder Executivo, de autoridade competente para zelar pela fiscalização da aplicação da lei e estabelecer normas e diretrizes para sua implementação, tal qual ocorre com a Autoridade Nacional de Proteção de Dados (ANPD), no âmbito da LGPD.

Algumas previsões são polêmicas, a exemplo da responsabilidade civil, que, na nova proposta, é objetiva para o operador de sistema de inteligência artificial de alto risco ou de risco excessivo que causar dano patrimonial, moral, individual ou coletivo (art. 27, §1º), em oposição à proposta inicial aprovada na Câmara dos Deputados, que previa responsabilidade subjetiva e foi objeto das discussões travadas pela comissão, diante da dificuldade de se comprovar a conduta culposa do desenvolvedor da tecnologia e por ser contrária à prevalência do regime de responsabilidade objetiva consagrado na doutrina, na jurisprudência, no Código Civil e na legislação consumerista.

Nos demais casos, ou seja, quando não envolver alto risco ou risco excessivo, a proposta prevê presunção de culpa do agente causador do dano, aplicando-se a inversão do ônus da prova em favor da vítima (art. 27, §2º).

Por fim, importante mencionar que a minuta proposta pela Comissão de Juristas, com o objetivo de fomentar a inovação, previu a possibilidade de que a entidade fiscalizadora autorize o ambiente regulatório experimental, denominado *Sandbox* Regulatório. Sobre o instituto, observa-se que, *in verbis*:

O objetivo do *Sandbox* Regulatório é criar um ambiente de negócios em que as empresas possam testar produtos e serviços inovadores sem as limitações e restrições impostas pelas regras vigentes, estimulando, dessa forma, o ambiente de inovação tecnológica, sem que se perca o controle sobre os novos negócios que estão surgindo, a fim de mitigar eventuais riscos às empresas, à economia e aos clientes. A regulação dos negócios baseados em soluções inovadoras e disruptivas não acompanha a velocidade de transformação desses negócios, então, essa ferramenta veio para auxiliar na adequação das regras à nova realidade imposta por essas empresas, possibilitando que regras sejam testadas em pequenos grupos antes que tenham alcance geral.¹⁷

O Projeto de Lei n. 21/2020 segue em tramitação no Senado Federal, e a expectativa é de que seja amplamente debatido antes de ser votado e de retornar à Câmara dos Deputados, diante da relevância da matéria que, como se viu, envolve diretamente a proteção a direitos e valores fundamentais e o futuro da inovação no Brasil.

CONCLUSÃO

A permeabilidade das técnicas de inteligência artificial em variados campos da sociedade é uma realidade que não pode ser ignorada. As novas tecnologias têm mostrado sua capacidade de

¹⁷ BRASIL. Tribunal de Contas da União. **Sandbox regulatório**. Disponível em: <https://portal.tcu.gov.br/data/files/86/75/3F/4B/669A38102DFE0FF7F18818A8/SandBox%20Regulatorio.pdf>. Acesso em: 14 mar. 2023.

revolucionar o mundo como o conhecemos e, caso bem empregadas, podem constituir ferramentas relevantes para a melhoria da qualidade de vida da população.

No âmbito da Administração Pública, como se viu, a utilização da inteligência artificial tem o potencial de propiciar a prestação de serviços públicos mais eficientes, de reduzir a burocracia e aprimorar o controle da atividade estatal.

No entanto, as tecnologias disruptivas também trazem consigo riscos a direitos e garantias fundamentais, que desafiam os operadores do direito a buscar regular a forma como se darão essas interações, sobretudo em razão da multidisciplinaridade que envolve o tema e da velocidade com que novas técnicas de inteligência artificial surgem.

Diante disso, o Brasil tem buscado, por meio do Projeto de Lei n. 21/2020, em trâmite no Senado Federal, estabelecer um marco normativo que estabeleça princípios, regras, diretrizes e fundamentos para regular o desenvolvimento e a aplicação da inteligência artificial no Brasil.

A iniciativa é bem-vinda e está em consonância com as tendências internacionais que visam estabelecer parâmetros para o uso ético das novas tecnologias.

No entanto, as previsões específicas do Projeto de Lei precisam ser largamente debatidas pela sociedade antes de aprovadas, tendo em vista que a regulação da inteligência artificial possui como principal desafio harmonizar a busca de segurança jurídica para os atingidos pelas ferramentas de inteligência artificial, induzindo boas práticas, com o incentivo à inovação, já que a criação de barreiras pode impactar no crescimento tecnológico do país.

Assim, espera-se que o Estado, além de atuar de forma a regular o uso das novas tecnologias, também se ocupe de implementar programas de fomento e incentivo à inovação, para que o Brasil não seja apenas consumidor de tecnologias criadas no exterior, mas também possa produzir novos mecanismos de inteligência artificial capazes de solucionar os principais desafios da sociedade brasileira.

REFERÊNCIAS

AGENCIA DE GOBIERNO ELECTRÓNICO Y SOCIEDAD DE LA INFORMACIÓN Y DEL CONOCIMIENTO (AGESIC). Principios Generales sobre Inteligencia Artificial para un Gobierno Digital.

Disponível em: <https://bit.ly/2Siy4tO>. Acesso em: 24 out. 2022.

BINENBOJM, Gustavo. Poder de polícia; ordenação regulação. Belo Horizonte: Fórum, 2020.

BRASIL. Conselho Nacional de Justiça. Resolução n. 332, de 21 de agosto de 2020. Disponível em:

<https://atos.cnj.jus.br/atos/detalhar/3429>. Acesso em: 6 mar. 2023.

BRASIL. Ministério da Ciência, Tecnologia e Inovação. Estratégia Brasileira de Inteligência Artificial (Ebia). Disponível em:

www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/inteligencia-artificial. Acesso em: 14 mar. 2023.

BRASIL. Senado Federal. Coordenação de Comissões Especiais, Temporárias e Parlamentares de Inquérito. Relatório final da Comissão de Juristas responsável por subsidiar a elaboração de Substitutivo sobre inteligência artificial instituída pelo Ato do Presidente do Senado n. 4, de 2022, aprovado em 1º de dezembro de 2022. Disponível em:

<https://legis.senado.leg.br/sdleg-getter/documento?dm=9221643&ts=1677521675945&disposition=inline>. Acesso em: 15 mar. 2023.

BRASIL. Tribunal de Contas da União. Sandbox regulatório. Disponível em:

<https://portal.tcu.gov.br/data/files/86/75/3F/4B/669A38102DFE0FF7F18818A8/SandBox%20Regulatorio.pdf>. Acesso em: 14 mar. 2023.

GIAMBIAGI, Fabio; ALÉM, Ana Claudia. Finanças públicas: teoria e prática no Brasil. Rio de Janeiro: Elsevier, 2011.

GORZONI, Paula. Inteligência artificial: riscos para direitos humanos e possíveis ações. Disponível em: <https://itsrio.org/wp-content/uploads/2019/03/Paula-Gorzoni.pdf>. Acesso em: 9 mar. 2023.

HEIKKILÄ, Melissa. A Quick Guide to the Most Important AI Law You've Never Heard of. Disponível em: www.technologyreview.com/2022/05/13/1052223/guide-ai-act-europe/?utm_campaign=site_visitor.unpaid.engagement&utm_medium=tr_social&utm_source=Twitter. Acesso em: 6 mar. 2023.

IVAR, A. et. al. Regulação de inteligência artificial no Brasil: policy paper. 2020. Disponível em: <https://bibliotecadigital.fgv.br/dspace/handle/10438/30078>. Acesso em: 24 out. 2022.

KELLER, Clara Iglesias; BAPTISTA, Patrícia. Por que, quando e como regular as novas tecnologias? Os desafios trazidos pelas inovações disruptivas. Revista de Direito Administrativo - RDA, ano 11, n. 273, set./dez. 2016. Disponível em: www.forumconhecimento.com.br/periodico/125/347/3395. Acesso em: 1º mar. 2023.

OCDE. Recommendation of the Council on Artificial Intelligence. Maio 2019. Disponível em: <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449>. Acesso em: 9 mar. 2023.

SCHWAB, Klaus. A quarta revolução industrial. Tradução de Daniel Moreira Miranda. São Paulo: Edipro, 2016.

DE LAS ACTUACIONES ADMINISTRATIVAS AUTOMATIZADAS A LA INTELIGENCIA ARTIFICIAL. UN BREVE APUNTE SOBRE SU IMPLEMENTACIÓN Y REGULACIÓN EN LA ADMINISTRACIÓN PÚBLICA ESPAÑOLA

Elisa Rocío Prados Pérez
Doctora en Derecho
Oficial Mayor Ayuntamiento Huelva

A MODO DE INTRODUCCIÓN

La Fundación del Español Urgente otorgó el título de palabra del año 2022 a la expresión “Inteligencia artificial”. Su selección se debía a su gran presencia en los medios de comunicación y al intenso debate social suscitado por los avances producidos en este ámbito y las consecuencias éticas derivadas del mismo.

Vivimos en un mundo de cambio, incertidumbre, mucha volatilidad, globalización, revolución tecnológica, lo que se ha llamado la “cuarta revolución industrial”, en la que el espacio y la tecnología digital se impone frente a lo analógico, que transforma nuestra sociedad y en consecuencia, nuestro sistema productivo.

La Inteligencia artificial es algo que forma parte de nuestra vida cotidiana. La utilizamos de forma habitual y sin darnos cuenta para buscar el camino más rápido para ir a la oficina, para traducir páginas web y también son asombrosas sus aplicaciones ligadas al desarrollo creativo de contenido artístico ya sea en el campo de la pintura o la música. 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.

Por su parte, la Administración se está enfrentando a un acelerado proceso de transformación que conlleva una nueva forma de entender las relaciones con los ciudadanos. La digitalización de la organización y de sus procedimientos de actuación nos sitúan ante un cambio de paradigma. En ese marco de transformación digital, la innovación tecnológica debe traducirse en innovación administrativa. Y esta innovación debe venir de la incorporación de la Inteligencia artificial que es una de las técnicas disruptivas llamadas a transformar no sólo a la sociedad sino también a nuestras administraciones.

En palabras del profesor SANCHEZ BRAVO “Las nuevas tecnologías digitales son ya una realidad. Entiendo que es vital hacer una reflexión acerca del papel que corresponde a las instancias públicas. “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 y para siempre todos los males que aquejan a la sociedad actual. Enfrente, los detractores del sistema. 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”.

Así las cosas, la administración del futuro no sólo será electrónica sino también inteligente y va a requerir de empleados públicos con unos perfiles profesionales bien diferentes de los actuales y a los que se les demandan una enorme flexibilidad y adaptabilidad orientado a la innovación. Si la administración está llamada a ser la

¹ Vid. SANCHEZ BRAVO, España Digital 2025. Estrategia Nacional de Inteligencia Nacional, p. 526

principal impulsora de los cambios, los ciudadanos tienen que ser también protagonistas de estos cambios.

LAS ACTUACIONES ADMINISTRATIVAS AUTOMATIZADAS: NUESTRO PUNTO DE PARTIDA

La entrada en vigor de las *Leyes 39/2015 y 40/2015, de 1 de octubre, del Procedimiento Administrativo Común y del Régimen Jurídico del Sector Público*, conllevó la necesaria realización de actuaciones encaminadas a la implantación del procedimiento administrativo tramitado por medios electrónicos.

Habida cuenta que los procedimientos de las Administraciones deben tramitarse electrónicamente y que deben adaptarse al Reglamento estatal de actuación y funcionamiento del sector público por medios electrónicos, aprobado por *Real Decreto 203/2021, de 30 de marzo* y que una de las manifestaciones más comunes para actos repetitivos y burocráticos son los que corresponden las actuaciones administrativas automatizadas.

En concreto, la citada *Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público (LRJSP)*, regula en su *artículo 40* los sistemas de identificación y firma de las Administraciones Públicas, estableciendo que podrán identificarse mediante el uso de un sello electrónico basado en un certificado electrónico reconocido o cualificado que reúna los requisitos exigidos por la legislación de firma electrónica.

Asimismo, el *artículo 41* de la citada *Ley 40/2015* define la actuación administrativa automatizada como cualquier acto o actuación realizada íntegramente a través de medios electrónicos por una Administración Pública en el marco de un procedimiento administrativo y en la que no haya intervenido de forma directa un empleado público.

A continuación, el *apartado 2 del artículo 41* recoge que en caso de actuación administrativa automatizada deberá establecerse previamente el órgano u órganos competentes, según los casos, para la definición de las especificaciones, programación, mantenimiento, supervisión y control de calidad y, en su caso, auditoría del sistema de información y de su código fuente, indicando el órgano que debe ser considerado responsable a efectos de impugnación.

Por otra parte, en el *artículo 42 de la Ley 40/2015* se establece que, en el ejercicio de la competencia en la actuación administrativa automatizada, cada Administración Pública podrá determinar los supuestos de utilización, como sistema de firma, del sello electrónico basado en certificado electrónico reconocido o cualificado que reúna los requisitos exigidos por la legislación de firma electrónica.

En la misma línea, el *artículo 20 del Real Decreto 203/2021, de 30 de marzo, que aprueba el Reglamento de actuación y funcionamiento del sector público por medios electrónicos* señala que en la tramitación administrativa autorizada de los procedimientos cada Administración Pública podrá determinar como sistema de firma electrónica el sello electrónico de órgano, permitiendo así, reducir los plazos y cargas de trabajo.

La actuación administrativa automatizada permite la realización de actuaciones administrativas sin necesidad de intervención por parte del personal de la Administración, agilizando con ello la gestión de los expedientes administrativos y garantizando la disponibilidad de los servicios electrónicos, al permitir su aplicación las 24 horas del día los 7 días de la semana.

A modo de ejemplo, con carácter general, pueden considerarse como actuaciones administrativas automatizadas a aquellas actuaciones y procedimientos que se encuentren incorporados en cada momento, con carácter automatizado, en la sede electrónica

de la Administración de que se trate y con carácter específico a determinadas actuaciones que pudieran ser, sin ánimo de ser exhaustivos, las que a continuación se describen: actos automáticos de impulso y de mero trámite del procedimiento electrónico, la emisión de certificado automático de inserción de anuncios en los tablones electrónicos, la expedición automática de recibos de presentación por el registro electrónico, el foliado automático de expedientes electrónicos, la emisión automática de copias auténticas de documentos electrónicos en los términos previstos en el *artículo 27 de la Ley 39/2015, de 1 de octubre, del Procedimiento Administrativo Común de las Administraciones Públicas (LPAC)*.

También al hilo de lo anterior, serían actuaciones administrativas automatizadas la digitalización de documentos en soporte papel o en otro soporte no electrónico en los términos del citado *artículo 27 de la LPAC*, las migraciones y cambios de formato automáticos, en particular, a efectos de archivo; las diligencias sobre documentos electrónicos o en soporte papel para la constancia de la fecha de la resolución o acuerdo en la que fueron aprobados. Asimismo, pueden realizarse a través de actuaciones administrativas automatizadas la firma de las resoluciones y notificaciones de los órganos colegiados y unipersonales, la fe pública en los contratos y convenios administrativos, la asignación de la numeración de las resoluciones y acuerdos y su inclusión en el Libro electrónico, las diligencias de apertura y cierre de los libros electrónicos, la expedición automática de recibos de presentación en los Registros electrónicos. Continuamos con otros trámites como la expedición de certificados y volantes de documentos del padrón municipal, histórico de movimientos, etc, la expedición electrónica de certificados e informes de datos individualizados de los padrones, matrículas, censos y demás documentos tributarios e ingresos públicos, la expedición electrónica de certificados en base a Registros públicos, comunicaciones a la ciudadanía, entre otros.

Trámites que permiten celeridad y rapidez en los procedimientos y en la formación de los expedientes siempre con las debidas garantías de tramitación y basadas en el cumplimiento de las características técnicas generales del sistema de firma y certificado aplicables; que serán conformes a lo establecido en el Esquema Nacional de Interoperabilidad en el ámbito de la Administración Electrónica, normativa concordante, en la propia política de firma y política de archivo en soporte electrónico de la Administración. En todo caso, la relación de los sistemas de obtención de sello electrónico utilizado, incluyendo las características de los certificados electrónicos y los prestadores que los expiden, será pública y accesible en la sede electrónica de la Administración.

Para la identificación y la autenticación del ejercicio de la competencia en la actuación administrativa automatizada, se utilizará con carácter general, un sistema de Código Seguro de Verificación (en adelante, CSV) vinculado al documento. El sistema generará el CSV mediante un algoritmo aleatorio y lo vincula al documento electrónico en que se materialice la actuación automatizada y a la aplicación firmante del mismo. El CSV solo podrá ser utilizado para autenticar aquellas actuaciones automatizadas que se hayan publicado en la sede electrónica. El sello electrónico garantizará la autenticidad y la integridad del documento emitido, a cuyo efecto contendrá un CSV con que se haya firmado, que podrá consultarse en la sede electrónica de la administración de que se trate. La validación del certificado podrá realizarse a través de los servicios ofrecidos y publicados por la web del prestador de certificados de certificación de confianza cualificado.

La actuación administrativa automatizada debe aprobarse por resolución administrativa y deberá publicarse en la Sede Electrónica de la Administración. Todo ello sin perjuicio de otras garantías adicionales de publicidad en el Boletín Oficial que corresponda. Contra la resolución aprobatoria de la actuación administrativa automatizada, al ser acto administrativo, que es

definitivo en vía administrativa, se podrá interponer, con carácter potestativo, recurso de reposición ante el órgano que lo ha dictado, en el plazo de un mes a contar desde la publicación en la sede electrónica. Alternativamente, se podrá interponer recurso contencioso-administrativo ante los Juzgados competentes de la Jurisdicción Contencioso-Administrativa.

Las Actuaciones Administrativas Automatizadas son una realidad que resulta actualmente de aplicación ordinaria en muchas de nuestras administraciones pero el final del camino, no es sólo eso sino mucho más, esto es, la utilización y el uso de la inteligencia artificial que está ya entre nosotros como veremos a continuación. En la presente aportación trataremos de reflexionar sobre los retos, oportunidades y posibilidades que plantea las diferentes herramientas y sistemas de Inteligencia artificial en la administración. Igualmente, realizaremos un breve apunte sobre los usos de la inteligencia artificial en la Administración y sus riesgos. Entendemos que es necesario profundizar en las limitaciones éticas que plantea su desarrollo para que facilite su crecimiento responsable en beneficio de toda la sociedad.

LA INTELIGENCIA ARTIFICIAL APLICADA A LA ADMINISTRACIÓN: UNA REALIDAD CADA VEZ MÁS CERCANA

Para abordar estas cuestiones, debemos señalar los tres estadios que concurren en la práctica diaria de nuestras Administraciones. De un lado, se ha producido una progresiva digitalización en un entorno de las nuevas tecnologías y de comunicación, que nos ha traído un inmenso tratamiento de datos y su debida trazabilidad. Junto a la digitalización, aparece la automatización cuya finalidad es sustituir al operador humano en las tareas burocráticas, repetitivas, instrumentales, como hemos visto anteriormente y que están reguladas en el art 41 de la Ley 40/2015. El tercer y definitivo paso en este proceso es el de la Inteligencia artificial (IA), cuya implementación está orientado a la toma de decisiones,

sustituyendo al ser humano. Estos tres estadios conviven, son complementarios pero a la vez independiente y todos son importante para apoyar la toma de decisiones en el ámbito de la administración pública.

Cuando hablamos de IA nos estamos refiriendo al desarrollo de sistemas y algoritmos que son capaces de realizar tareas y de aprender mediante un entrenamiento basado en incorporar grandes conjuntos de datos para adquirir con ello nuevas y mejores capacidades. A partir de los datos en poder de las Administraciones, de los algoritmos y del aumento de la potencia computacional de los ordenadores y de los sistemas de almacenamiento de datos, la Inteligencia artificial puede cumplir un relevante papel en el seno de la Administración.

De esta manera, a través de la inteligencia artificial los ordenadores son capaces de realizar acciones que serían consideradas como “inteligentes” si fuesen desarrolladas por personas. Se trata, en realidad, de emular las facultades intelectuales humanas en máquinas para que éstas puedan realizar tareas propias de los seres humanos.

Dentro de la Inteligencia Artificial es importante tener en cuenta elementos específicos como son los algoritmos y los datos. Siguiendo a PONCE SOLÉ utilizando el símil en la cocina de la Inteligencia Artificial, los primeros, serían las recetas y los segundos, los ingredientes². Así es. Los algoritmos se define como “el procedimiento para encontrar la solución a un problema mediante la reducción del mismo a un conjunto de reglas”. En la actualidad, los algoritmos que tienen un mayor potencial son aquellos que permiten el aprendizaje automático (*machine learning*) a través del que los ordenadores adquieren la capacidad de aprender de los datos y la experiencia que les permiten tomar

² Vid. PONCE SOLÉ, J ., “Inteligencia artificial, derecho administrativo y reserva de humanidad: algoritmos y procedimiento administrativo debido tecnológico” en Revista General de Derecho Administrativo núm 50, Iustel, enero 2019.

decisiones por sí mismos. A partir de los datos disponibles, el algoritmo identifica patrones y correlaciones entre los datos que son utilizados posteriormente para generar un modelo cuya calidad es evaluada estadísticamente y, en su caso, mejorada a partir de los resultados que se van obteniendo. Progresando, un estadio más avanzado de inteligencia artificial es el aprendizaje profundo (*deep learning*) que se basa en el aprendizaje de complejas redes neuronales, que emulan el funcionamiento del cerebro humano, a partir del uso de ingentes cantidades de datos.

A este respecto, debemos plantearnos la naturaleza jurídica de un algoritmo, ¿es una norma jurídica, es un acto? De este modo, en la actual doctrina administrativa española existe un candente debate acerca de su naturaleza jurídica, existiendo una línea doctrinal defensora de su carácter reglamentario, y otra, en oposición a la primera, que aboga por su enmarcación como instrumento aplicativo del Derecho, pero no su consideración como norma que innove el ordenamiento jurídico, pero debemos tener presente que no todos los algoritmos que utiliza la Administración pública producen efectos jurídicos. Por otro lado, sobre este particular entendemos su complejidad porque muchos de los algoritmos no se agotan con su cumplimiento o uso y algunos de ellos, los que utilizan el aprendizaje automático, pueden llegar a innovar el ordenamiento jurídico al incorporar criterios que no están explícitamente previstos en la norma. En cualquier caso, debemos advertir que, el procedimiento de elaboración de los algoritmos dista mucho de cumplir con trámites propios del procedimiento de elaboración de los reglamentos y, en particular, su publicación.

Esto nos debe hacer reflexionar sobre los algoritmos y sobre lo relevante que es la claridad y la transparencia en su diseño y en las instrucciones de su programación, que permitan el aprendizaje automático. Máxima trazabilidad en todo lo referente al modelo matemático de predicciones, su código fuente y sus características y para el que la legislación señala que ha de ser abierto. A este respecto uno de los riesgos que a menudo se han planteado es el

de caja negra, o de la opacidad del algoritmo, esto debe proscribirse.

Como hemos señalado anteriormente, para poder llevar a cabo todos estos procesos, el aprendizaje automático requiere grandes cantidades de datos (big data), datos masivos que se identifican con tres características: volumen, variedad y velocidad (las tres v). Su recopilación, ordenación, estructura, análisis e interpretación deviene esencial. A modo de ejemplo, los Ayuntamientos de Barcelona y Vitoria tiene ya una Oficina del dato porque los datos en la hora presente han adquirido una importancia capital. Han de servir para las diferentes áreas y servicios administrativos así como han de estar a disposición de la sociedad civil en general. Esto enlaza con la idea de gobierno abierto y con la idea de crear valor en torno al dato.

Parece indiscutida la posibilidad de usar la Inteligencia Artificial en aquellos expedientes donde se producen decisiones automatizadas en relación con potestades regladas pero hemos de manifestar mayor precaución para admitir decisiones automatizadas en ámbitos donde impera la discrecionalidad administrativa. La legislación alemana de procedimiento administrativo ha prohibido el uso de la inteligencia artificial en las decisiones discrecionales y en la misma línea el *art 44.2 de la Ley catalana 26/2010, de 3 de agosto, de régimen jurídico y de procedimiento de las administraciones públicas* señala que “sólo son susceptibles de actuación administrativa automatizada los actos que puedan adoptarse con una programación basada en criterios y parámetros objetivos”. Es, por ello, que a sensu contrario, lo subjetivo, determinante de la ponderación discrecional está vedada en la legislación catalana.

Consideramos que en la Administración pública, la decisión de poner en funcionamiento un sistema de inteligencia artificial debe ser adoptado formalmente mediante un acto administrativo, una exigencia y condición inexcusables para facilitar la adecuada

defensa de los derechos e intereses legítimos ya que, en definitiva, de no ser así, se estaría dificultando una eventual impugnación por parte de los ciudadanos destinatarios. Asimismo, la intención de utilizar inteligencia artificial en un determinado supuesto debería hacerse con previa participación ciudadana, contar con el trámite de información pública previsto en el *artículo 83 LPAC*, que favorezca la transparencia. Como principio de actuación de las administraciones públicas, la transparencia debe guiar el uso de la inteligencia artificial en las administraciones públicas, debiendo ser objeto de publicidad todos esos extremos de cómo se adoptan estas decisiones públicas.

Igualmente, deviene esencial en la adopción de mecanismos de IA la obligación de motivación como uno de los pilares del derecho a la buena administración que lo justifica a fin de dar respuesta y ofrecer un mejor control de esa decisión administrativa, no solo en vía administrativa, sino también judicial e incluso por la Agencia Española de protección de Datos. El incumplimiento de estas cautelas podría tener consecuencias, como no puede ser de otra manera, desde la perspectiva de la invalidez de los actos administrativos o, en su caso, de la responsabilidad patrimonial de la Administración Pública asumiendo las consecuencias de un funcionamiento anormal de los servicios públicos en los términos del *art 32 LRJSP*.

LA NECESIDAD DE UN MARCO NORMATIVO Y ÉTICO ADECUADO: EL LIDERAZGO EUROPEO EN LA TRANSICIÓN DIGITAL

Aunque la tecnología y la sociedad siempre se adelantan al Derecho, es necesario un marco normativo y ético adecuado que encuadre la Inteligencia Artificial. Desde las más diversas instancias se están realizando esfuerzos, entre los que merece destacarse la Recomendación de la UNESCO sobre la ética de la

Inteligencia Artificial y la Recomendación sobre Inteligencia Artificial de la OCDE a la que se han adherido multitud de países³.

Pero es respecto a la Inteligencia Artificial, donde la Unión Europea está consolidando un enfoque propio. Esta política regulatoria podría suponer una desventaja competitiva frente a otros lugares en los que no prevén esas cautelas y siendo conscientes de ello, las instituciones europeas aspiran a que, con su ejemplo, se logre un efecto emulador en otros países.

Sobre Inteligencia Artificial, la Comisión Europea ha aprobado dos relevantes documentos: la *Comunicación de la Comisión Europea sobre Inteligencia artificial para Europa COM (2018) 237, de 25 de abril*; y el Plan coordinado sobre la inteligencia artificial, *Comunicación de la Comisión Europea COM (2018) 795, de 7 de diciembre*. Después, aprobó la *Comunicación COM (2019) 168, de 8 de abril*, denominada Generar confianza en la inteligencia artificial centrada en el ser humano. Seguidamente, debemos destacar igualmente el *Libro Blanco de la UE sobre la inteligencia artificial*: un enfoque europeo orientado a la excelencia y a la confianza, en el que se establecen las líneas maestras de una futura regulación en la materia íntimamente conectado con la Estrategia europea de datos, *Comunicación de la Comisión COM (2020) 66 final, de 19 de febrero*. Igualmente, el Parlamento Europeo ha impulsado varias iniciativas, entre las que destacamos el *Marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas* y la *Resolución sobre Inteligencia artificial: cuestiones de interpretación y de aplicación del Derecho internacional*. Asimismo, debemos hacer mención al *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*.

³ Recomendación sobre la Ética de la Inteligencia Artificial, adoptada por la Conferencia General de la UNESCO, reunida en París del 9 al 24 de noviembre de 2021, en su 41ª reunión y Recomendación del Consejo sobre Inteligencia Artificial, de 22 de mayo de 2019.

Vemos que, desde la Unión Europea se está realizando un esfuerzo de ir adaptando el concepto de IA a la nueva realidad tecnológica y social. Por último, en lo más reciente, destacamos la “Propuesta de Reglamento del Parlamento Europeo y el Consejo estableciendo reglas armonizadas sobre inteligencia artificial”. Su rango normativo equivale al de una Ley europea y por tanto resulta de especial interés dado su carácter directo y vinculante tanto para los Estados miembros como para la ciudadanía logrando así un mayor grado de uniformidad en el régimen sustantivo y en su aplicación práctica. Se trata de lograr un marco normativo armonizado en toda la Unión, que evite la fragmentación del mercado único; que las exigencias impuestas a los desarrollos de la IA sean las mismas en toda la UE, permitiendo así la libre circulación de la tecnología. A tales efectos, la propuesta de Reglamento establece un relevante papel para la Junta europea de Inteligencia Artificial, compuesta por representantes de los Estados miembros y de la Comisión, así como para el Supervisor Europeo de protección de datos. Conviene puntualizar que esta propuesta de regulación adquiere la forma de Reglamento, se aplica de forma inmediata, sin necesidad de acto alguno de transposición al Derecho interno.

A la espera de la aprobación del Reglamento europeo, en España, se ha publicado un *Libro Blanco y en la Ley de Presupuestos Generales de 2022*, que ha incluido una dotación para una nueva agencia que supervisará las aplicaciones que cuenten con inteligencia artificial. Y, recientemente, se han incorporado unas mínimas pautas para evitar sesgos, facilitar la transparencia y la rendición de cuentas a través del *art. 23 de la Ley 15/2022, de 12 de julio, de igualdad de trato y no discriminación* que señala “ 1. En el marco de la Estrategia Nacional de Inteligencia Artificial, de la Carta de Derechos Digitales y de las iniciativas europeas en torno a la Inteligencia Artificial, las administraciones públicas favorecerán la puesta en marcha de mecanismos para que los algoritmos involucrados en la toma de decisiones que se utilicen en las administraciones públicas tengan en cuenta criterios de minimización de sesgos, transparencia y rendición de cuentas,

siempre que sea factible técnicamente. En estos mecanismos se incluirán su diseño y datos de entrenamiento, y abordarán su potencial impacto discriminatorio. Para lograr este fin, se promoverá la realización de evaluaciones de impacto que determinen el posible sesgo discriminatorio. 2. Las administraciones públicas, en el marco de sus competencias en el ámbito de los algoritmos involucrados en procesos de toma de decisiones, priorizarán la transparencia en el diseño y la implementación y la capacidad de interpretación de las decisiones adoptadas por los mismos. 3. Las administraciones públicas y las empresas promoverán el uso de una Inteligencia Artificial ética, confiable y respetuosa con los derechos fundamentales, siguiendo especialmente las recomendaciones de la Unión Europea en este sentido. 4. Se promoverá un sello de calidad de los algoritmos”

A nivel nacional, debemos hacer referencia al trabajo que desde el Gobierno de España, en concreto, desde la Secretaría de Estado de Digitalización e Inteligencia Artificial, se ha venido realizando fundamentalmente en los últimos años. *La Estrategia Nacional de Inteligencia Artificial (ENIA) de 2 de diciembre de 2020 y la Carta de Derechos digitales* vienen a constituir los dos pilares programáticos sobre los que pretende asentarse la planificación de la IA en España. Asimismo, es fundamental tener en cuenta la Estrategia Española de I+D+I en Inteligencia artificial del Ministerio de Ciencia, Innovación y Universidades.

En este sentido debemos subrayar aunque sea someramente el *Plan de Digitalización de las Administraciones Públicas (PDAP)*, con el horizonte temporal 2025, pretende un salto decisivo en la mejora de la eficacia y eficiencia de la Administración Pública. Sus tres objetivos fundamentales pueden agruparse en:

a) Servicios digitales, accesibles, eficientes, seguros y fiables que permitan 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 ciudadanos 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: desarrollar servicios, activos e infraestructuras comunes que permitan a todas las Administraciones Públicas sumarse a la revolución tecnológica que está suponiendo la irrupción de nuevos habilitadores tecnológicos como pueden ser la Inteligencia Artificial o la tecnología de analítica de datos.

Por su parte, también hay que subrayar que el *14 de julio de 2021 se presentó la Carta de Derechos Digitales*, 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”.

No hemos de soslayar que cuando estemos ante decisiones automatizadas o con intervención de Inteligencia Artificial en su adopción se deberían desarrollar evaluaciones de impacto de protección de datos. A este tipo de evaluación se refiere el *artículo 35 del Reglamento General de Datos de la Unión Europea y también el artículo 28 de la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales*. A este respecto que el *artículo 22 del Reglamento europeo de protección de datos* resulta suficientemente expresivo al establecer que cualquier ciudadano tiene el derecho “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”.

Sobre este particular merece destacarse la sentencia del *Juzgado Central número 8 de 30 de diciembre de 2021*. Este caso trae causa de la petición efectuada por la Fundación Civio a través del Portal de Transparencia del Ministerio de transición ecológica donde se solicitaba las especificaciones técnicas y el código fuente del programa BOSCO (acrónimo de BOno Social Comercializadora). Se trata de programa que regula el bono social que reduce el pago de la factura de la comercializadora de energía eléctrica que atiende a los consumidores que se encuentran en una situación vulnerable o en riesgo de pobreza energética. De esta manera, se concede, entre otros, a aquellos que tuvieran una reducida potencia contratada así como aquellos pensionistas que estuvieran recibiendo la cuantía mínima. Ante el silencio, la Fundación se dirigió al Consejo de Transparencia y Buen Gobierno que estimó parcialmente la petición obligando a que aporte las especificaciones técnicas, pruebas y otra información sobre el funcionamiento del programa, no así el código fuente solicitado. Frente a esta resolución del Consejo de Transparencia se interpuso recurso contencioso-administrativo solicitando el código fuente, dicho recurso fue desestimado por la sentencia antes dicha señalando que siempre en caso de que el beneficiario se le denegara el Bono puede impugnarlo tanto en la vía administrativa como judicial y que no existe ninguna norma que imponga a la administración el desarrollo de aplicaciones con fuentes abiertas ni la adquisición de software libre⁴.

⁴ Vid. La Resolución del Consejo de Transparencia y Buen Gobierno 791/2018, de 18 de febrero de 2019 y la Sentencia del Juzgado Central núm. 8 de 30 de diciembre de 2021 y que está disponible en la página del Consejo de Transparencia:

https://www.consejodetransparencia.es/ct_Home/Actividad/recursos_jurisprudencia/Recursos_AGE/2019/128_particular_35.html

La transparencia es uno de los principios de actuación de las administraciones públicas que ha adquirido un mayor reconocimiento y que está teniendo un mayor impacto en el proceso de transformación de las administraciones públicas en los últimos años. Se puede canalizar a través de distintos mecanismos como la difusión proactiva de la información a través de Portales de transparencia o de otro lado, a través del ejercicio del derecho de acceso a la información pública. La transparencia consiste en el conocimiento por parte de la ciudadanía de lo que sucede en el seno de las Administraciones públicas, sobre las decisiones que toman, por qué las toman y cómo las toman. A este respecto, diversas decisiones de la GAIP catalana han concedido derecho de acceso al algoritmo, a modo de ejemplo, la *Resolución 200/2017, de 21 de junio*. La opacidad debe combatirse y en especial, para aquellas aplicaciones en las que los derechos fundamentales de las personas puedan resultar trasgredidos a consecuencia de esa falta de transparencia.

REFLEXIONES FINALES

A lo largo de nuestra exposición se ha evidenciado que la inteligencia artificial está suscitando intensos debates. A través de estas líneas hemos planteado brevemente algunos de los principales retos y problemas que el uso de la inteligencia artificial genera en el Derecho y en las Administraciones públicas.

La incorporación de la inteligencia artificial en la tramitación del procedimiento administrativo y, en general, en los procesos de toma de decisiones va a revolucionar la manera de actuar de las administraciones públicas alcanzando cada vez más un papel importante.

El desarrollo de sistemas de inteligencia artificial plantea enormes expectativas de desarrollo económico y social, permitiendo sensibles mejoras de prestaciones y servicios en el sector público aunque su uso suscita grandes dilemas jurídicos y éticos.

Desde principios de 2018 con la aprobación por la Comisión Europea de la *Comunicación de Estrategia sobre Inteligencia Artificial para Europa*, la Unión Europea ha venido trabajando intensamente en esta materia, teniendo como colofón la aprobación de una propuesta de Reglamento sobre IA que pretende convertirse en la primera normativa vinculante en esta materia. Una ambiciosa propuesta que promueve un marco regulatorio uniforme en todos los Estados miembros que, de un lado, impulse la seguridad jurídica y de otro, preserve los derechos fundamentales de los ciudadanos en una sociedad democrática. En este sentido presenta una perspectiva ética y antropocéntrica de la IA y focaliza su articulado en función del riesgo que los sistemas IA suponen para la persona, en lo que ha venido a ser una juridificación de los valores éticos centrados en el ser humano.

Así las cosas, para avanzar y valorar adecuadamente los diferentes impactos, contempla un debate interdisciplinar sobre el papel que la inteligencia artificial debe tener y orientar las decisiones que tomen al respecto. Considerando siempre como referencia y principio de actuación, la transparencia; que debe guiar el uso de la inteligencia artificial en las administraciones públicas para evitar efectos no deseados y perniciosos y evitar argumentos como el que se suscitó en su momento con *Matrix* (1999), película que abordaba una crítica a un mundo donde la IA había tomado el control de la vida y en la que los humanos se rebelaban contra ese dominio.

En el entorno de la administración son numerosos los ejemplos de aplicación de IA entre otros y en el ámbito eminentemente local, la comprobación de la calidad de aguas destinadas al abastecimiento de la población, las alertas de fugas en la canalización o en los filtros de la depuradora, la organización de las rutas y horarios de transportes públicos, los procesos de reciclado de residuos, el diseño de las pruebas de selección de personal, la utilización de chatbots para dar respuesta a las preguntas formuladas por la

ciudadanía y el uso de los drones para seguridad ciudadana, entre otros.

Dentro del estudio de la aplicación de la Inteligencia Artificial constituye una parte muy importante la debida conciliación y respeto de la protección de datos de los ciudadanos, ya que algunos de los datos que han de facilitar los usuarios serán, lógicamente, “personales” que han de protegerse y otros serán “sensibles” en el sentido establecido por el Reglamento europeo y la Ley Orgánica de Protección de Datos Personales española. Este es un tema muy importante que debe salvaguardarse.

Todo lo anterior debe contextualizarse en el derecho que todos tenemos a una buena administración contenido en el artículo 41 de la *Carta de Derechos Fundamentales de la Unión Europea* de 7 de diciembre de 2000 y concretado, después, en el *Código Europeo de Buena Conducta Administrativa* de 2005.

Ante estos retos, la cuestión principal que debe abordarse es 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. En nuestras manos está.

REFERENCIAS BIBLIOGRÁFICAS

- BOIX PALOP, A., “Los algoritmos son reglamentos”, en *Revista de Derecho Público: teoría y método*, n.1, 2020, pp. 223-269.
- CERRILLO I MARTINEZ, A., “El derecho para una inteligencia artificial centrada en el ser humano y al servicio de las instituciones”, *Revista de Internet, Derecho y Política* núm 30, p.1 y ss.
- DE MIGUEL ASENSIO, P., *Propuesta de Reglamento sobre inteligencia artificial*, en *La Ley Unión Europea*, n. 92, 2021.
- GAMERO, E, *Cumplimiento normativo de desarrollos de inteligencia artificial para la toma de decisiones administrativas* en *Diario La Ley* n.º 50, 2021.
- HUERGO LORA, A:

“Una aproximación a los algoritmos desde el derecho administrativo” en el libro colectivo *La regulación de los algoritmos*, Thomson Reuters Aranzadi, Cizur Menor, 2020, pp. 23-87.

“Administraciones públicas e inteligencia artificial ¿más o menos discrecionalidad?”, *El Cronista*, núm. 96-97, 2021, p. 78 y ss.

PONCE SOLÉ, J., “Inteligencia artificial, derecho administrativo y reserva de humanidad: algoritmos y procedimiento administrativo debido tecnológico” en revista general de Derecho Administrativo núm 50, Iustel, enero 2019.

SÁNCHEZ BRAVO, A

Derechos humanos y protección de datos personales en el Siglo XXI: homenaje a Cinta Castillo Jiménez, Punto Rojo Libros, Sevilla, 2014.

Derecho, inteligencia artificial y nuevos entornos digitales, Punto Rojo Libros, Sevilla, 2020.

VALERO TORRIJOS, J., “Las garantías jurídicas de la inteligencia artificial en la actividad administrativa desde la perspectiva de la buena administración” en la revista *Catalana de Dret Public* núm 58, 2019.

UM CONTROLE SOCIAL INOVADOR SOBRE O CONTROLE EXTERNO: MÉTODO PARA AFERIÇÃO DE DECISÕES DOS TRIBUNAIS DE CONTAS

Geraldo Costa da Camino.

Pós-doutorando em Controle Externo na Escola de Artes, Ciências e Humanidades da Universidade de São Paulo (EACH/USP). Procurador do Ministério Público junto ao Tribunal de Contas do Estado do Rio Grande do Sul.

O controle externo da Administração Pública, de titularidade do Poder Legislativo, é exercido com o auxílio - e através de competências próprias - do Tribunal de Contas. Órgãos de extração constitucional, as cortes de contas - que não integram os poderes, nem a eles se submetem - não formam propriamente um sistema, uma vez que, com seu viés marcadamente federativo, carecem de instâncias recursais exógenas e de uma lei orgânica nacional que lhes dê uniformidade processual, sendo regulamentadas localmente, nas respectivas esferas.

Talvez pela diversidade daí decorrente, quiçá pela especificidade da matéria que tratam, os tribunais de contas são superficialmente conhecidos, inclusive por considerável parcela da comunidade jurídica, e pouco se submetem ao escrutínio do controle social. Esse déficit de accountability pode ser enfrentado ao se “exigir [...] motivações mais extensas e didáticas ou requerer a utilização de procedimentos participatórios.” (JORDÃO, 2016, p. 244-245).

Identifica-se, a partir desse quadro, a necessidade de, especialmente no contexto das inovações tecnológicas que dão substrato às smart cities, avançar-se na interação entre Estado e Sociedade também no campo do controle externo. Como doutrinou Ricardo Lobo Torres, “só o controle [...] legitima o

tributo” (TORRES, 2008, p. 511). Se a importância do controle é inquestionável, sua legitimidade é impositiva, devendo ser aferida não somente no meio estatal, mas pela sociedade a quem serve o controle, e a quem se destina o objeto controlado.

Dromi (2005, p. 73) afirma que é preciso, ao defender a inserção recíproca do controle na sociedade e da sociedade no controle, cuidado com a coisa pública. Não basta, para tanto, estimular a sociedade civil a exercer o controle social. É preciso dotá-la de (1) conhecimento da atividade do controle externo e de (2) ferramentas tecnológicas para levá-lo a termo. Adaptando a sentença de Chamayou (2020, p. 22) sobre a governabilidade, pode-se dizer que a controlabilidade é uma capacidade composta da disposição de ser controlado e da aptidão para controlar.

Os mecanismos sociais, como destacam Smulovitz e Peruzzotti (2003, p. 315), podem influenciar o desempenho dos órgãos de controle horizontal (como os tribunais de contas), organizando-se em bases populares e servindo de “guardiões dos guardiões”. Para que isso ocorra, todavia, a participação desses atores não pode se restringir, como observa Barros (2019, p. 209), a inserções eventuais no processo de controle, despidas de protagonismo e levadas a termo de forma periférica.

Mais do que isso, esse controle precisa questionar a “eficácia procedimental dos órgãos de fiscalização dos gastos públicos” (BLIACHERIENE, TEIXEIRA e AZEVEDO, 2015, p. 121). Ainda mais importante se torna a ação do controle social ao se tomar a lição de Speck (2000, p. 16), no sentido de que as instituições regulatórias intermediárias, como as cortes de contas, são dotadas de alto grau de autonomia e não se sujeitam ao processo de controle popular, como os poderes Legislativo e Executivo.

Nesse sentido, é pertinente, por um lado, empreender pesquisa sobre a legitimidade da atuação do órgão de controle externo, verificando a coerência (inclusive a partir da cultura de

precedentes em voga no âmbito judicial) e a fundamentação de seus julgados (em atenção ao princípio da persuasão racional). A previsibilidade das decisões deve ser um atributo dos tribunais de contas, que não podem ser autorreferenciais, na acepção adotada por Luhman (1997, p. 51), da produção de resultados distintos a partir dos mesmos inputs.

O controle social de que se trata não deve afetar a liberdade de julgar, que é uma prerrogativa dos magistrados de contas, senão “refletir sobre a necessidade de procedimentos e parâmetros para a atividade jurisdicional que, como qualquer outra função atribuída a um estado de direito, não pode ser exercida arbitrariamente.” (RODRIGUEZ, 2013, p. 149). Muito menos chegar ao paroxismo cogitado por Han (2017, p. 110), ao advertir que “o controle total aniquila a liberdade de ação e leva, em última instância, à uniformização.”. Deve, entretanto, ser próximo - e indutor - do controle externo 4.0 que se anuncia, inclusive em relação aos próprios atos, como bem doutrinam Cunda e Ramos (2021, p. 215).

Ademais, não se pode prescindir de sindicar acerca de eventual ingerência política nas razões de decidir, em possível conexão com a forma de investidura de seus membros. Dentre tantos autores, Medauar (2014, p. 155) critica a verificação meramente formal de requisitos constitucionais, como os notórios conhecimentos, para essas indicações, comumente gestadas a partir de acordos políticos.

Podem resultar daí, como constata Lopes Jr. (2007, p. 109), prejuízos à eficiência e à legitimidade da atividade judicante dos tribunais, principalmente em face da chance de “interferência fundamental dos controlados [...] sobre os controladores” (BLIACHERIENE, 2016, p. 209).

Há um prognóstico de que “a inteligência artificial migrará para as posições de poder, responsabilidade e prestação de contas”

(SCHWAB e DAVIS, 2019, p. 181), abrindo possibilidade de atuação ainda mais efetiva – e essencial – para o controle social.

Impõe-se, portanto, buscar as experiências passadas e presentes no campo do controle social (observatórios, conselhos etc.), bem como as ferramentas de TI disponíveis para otimizá-las (em termos de dados abertos e inteligência artificial, por exemplo), no expansivo cenário das smart cities, inclusive no direito comparado. O controle social deve ser exercido com mais vigor, fazendo uso das novas tecnologias, conforme registram corretamente Ribeiro, Bliacheriene e Santana (2019, p. 39).

Interessante exemplo traz Regatieri (2020, p. 374), discorrendo sobre avaliação de políticas públicas no Reino Unido e no Canadá. Nesses países, essa tarefa pode ser exercida por gestores envolvidos no produto a ser aferido, mas sem prejuízo de que outros órgãos efetuem o controle de qualidade quanto à robustez das avaliações realizadas. Parece cabível uma analogia com o papel imaginado para o controle social em relação ao controle externo.

A indexação dos acórdãos dos processos de contas, a disponibilização prévia e tempestiva das pautas de julgamento, as bases de dados de precedentes, as informações públicas sobre julgadores e gestores, dentre outros requisitos, podem permitir que o controle social não seja meramente formal, senão avance na materialidade das decisões, e no questionamento sobre as razões de decidir, evidentes ou não.

É como entendem Lino e Aquino, em estudo de caso, para os quais “[m]esmo quando irregularidades gravíssimas são apontadas pelo corpo instrutivo, relatores com raízes político-partidárias tendem a sugerir a aprovação das contas municipais de seus correligionários [...] devido à limitação da participação social”. Acredita-se que o aperfeiçoamento da transparência e da segurança jurídica nas

ações de controle, com políticas de acompanhamento, impactará positivamente no combate à corrupção (RECH e ZORTÉA, 2020, p. 54).

Entretanto, não se desconhece o risco, como detectou Neves (2019, p. 49), dada a potencial influência política na ação do controle, de que a abordagem científica da matéria se faça em ambiente refratário a críticas. Para mitigá-lo, o pesquisador deve ser capaz de demonstrar a seu público-alvo a neutralidade do estudo empreendido, e os benefícios de seus produtos, não só, de forma ampla, à sociedade, senão, estritamente, à própria atividade dos órgãos analisados. É o que prega Speck (2013, p. 99), para quem os tribunais de contas só teriam a ganhar em termos de legitimidade tendo a sociedade como aliada, o que reverteria também em prol da eficácia de sua atividade.

Assim, justifica-se a pesquisa pela necessidade de ser suprida a ausência de um controle social efetivo, sistemático e inovador sobre o controle externo exercido pelos tribunais de contas. Se implementado, com o uso de novas tecnologias, esse controle propiciará, a curto prazo, a melhoria da atuação das cortes, em face do escrutínio de suas decisões. Corolário disso, a médio prazo, as políticas públicas submetidas ao controle externo serão otimizadas, mercê do incremento da legitimidade dos órgãos controladores.

O problema de pesquisa, portanto, consiste na possibilidade - ou não - de se inovar no controle social sobre o controle externo, com o uso de novas tecnologias capazes de rendê-lo efetivo, identificando conflitos de interesse, desconsideração de apontamentos de auditoria e incoerência com os próprios precedentes. A pergunta dele decorrente: é possível avançar no rumo da materialidade de um controle social efetivo e inovador sobre o controle externo realizado pelos tribunais de contas?

Como? O objetivo geral da pesquisa não se restringe ao aspecto prospectivo. Mais do que isso, tem ela intuito propositivo, visando a apresentar uma alternativas para a resolução do problema colocado.

Desta forma, a pesquisa tem por objetivo geral o de, a partir da análise das decisões dos tribunais de contas, desenvolver um método capaz de propiciar um controle social inovador sobre o controle externo, com parâmetros seguros e previsibilidade de resultados.

Decompondo o objetivo geral acima, tem por objetivos específicos, os de (1) analisar a estrutura das decisões dos tribunais de contas a fim de identificar e categorizar seus diferentes componentes em relação aos tipos de dados de entrada; (2) a partir da categorização do item (1), identificar características mais apropriadas para indexação e aferições intrínsecas (em cotejo com manifestações da auditoria e do Ministério Público de Contas no mesmo processo) e extrínsecas (levando em contas precedentes do próprio e/ou de outros tribunais) da adequação das decisões dos tribunais de contas; (3) propor, a partir de interlocuções com essas cortes (por meio de seus Marcos de Medição) e observatórios de Tribunais de Contas, um arcabouço para processamento das características extraídas no item (2); e (4) definir método para predição de resultados de julgamentos, capaz de servir de base para soluções tecnológicas úteis ao controle social sobre o controle externo.

Assim, o resultado esperado da pesquisa é um método de aferição da qualidade das decisões dos tribunais de contas que possa servir de insumo para a implementação de soluções com uso de novas tecnologias para o controle social sobre o controle externo da Administração Pública.

REFERÊNCIAS

- ALVES, F. D. Controle social de políticas públicas. Santa Cruz do Sul: Estudos de Direito, 2018.
- BARROS, L. M. A. Compliance e controle social do setor público: auditorias democráticas. São Paulo: Quartier Latin, 2019.
- BLIACHERIENE, A. C. Controle da eficiência do gasto orçamentário. Belo Horizonte: Editora Fórum, 2016.
- BLIACHERIENE, A. C.; TEIXEIRA, B. C.; AZEVEDO, D. Q. F. Teoria do Desenvolvimento e as perspectivas para um controle popular das contas públicas. Revista de Informação Legislativa, v. 205, p. 107-126, 2015.
- CHAMAYOU, G. A sociedade ingovernável: Uma genealogia do liberalismo autoritário. São Paulo: Ubu, 2020.
- CONTI, J. M. Direito financeiro na Constituição de 1988. São Paulo: Oliveira Mendes, 1998.
- CUNDA, D. Z. G.; RAMOS, L. A. Ciber@administração pública e Controle 4.0: seus desafios em tempos de pandemia do coronavírus, e a transparência ampliada (para além de translúcida). In: LIMA, L. H.; GODINHO, H. H. A. M.; SARQUIS, A. M. F. Os desafios do controle externo diante da pandemia da covid-19: estudos de ministros e conselheiros substitutos dos Tribunais de Contas. Belo Horizonte: Fórum, 2020. p. 207-242.
- DA CAMINO, G. C. A investidura no Tribunal de Contas. Belo Horizonte: Fórum, 2021.
- DA CAMINO, G. C. República como responsabilidade: o conteúdo jurídico do princípio republicano na Constituição brasileira. 2021. 285 p. Tese (Doutorado em Direito) – Faculdade de Direito, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2021.
- FONSECA, F.; GUEDES, A. M. Controle social da Administração Pública. São Paulo: UNESP, 2008.
- FRAZÃO, A.; MULHOLLAND, C. (coord.). Inteligência artificial e Direito: Ética, regulação e responsabilidade. 2. ed. rev., ampl. e atual. São Paulo: Revista dos Tribunais, 2020.
- GIL, A. C. Métodos e Técnicas de Pesquisa Social. 7. ed. São Paulo: Atlas, 2019.
- HAN, B. Sociedade da transparência. Petrópolis-RJ: Vozes, 2017.
- JORDÃO, E. Controle judicial de uma administração complexa: a experiência estrangeira na adaptação da intensidade do controle. São Paulo: Malheiros: SBDDP, 2016.

LINO, A. F.; AQUINO, A. C. B. Práticas não adequadas nos tribunais de contas. *Revista de Administração Pública*, 54(2), 220-242, 2020.
Disponível em:
<https://www.scielo.br/j/rap/a/S5G5y97GWRyB64B3PDmwKjk/?lang=pt>. Acesso em: 22 jul. 2021.

LOPES JUNIOR, E. M. A judicialização da política no Brasil e o TCU. Rio de Janeiro: FGV, 2007.

LUHMAN, N. Novos desenvolvimentos na teoria dos sistemas. In: NEVES, C. E. B.; SAMIOS, E. M. B. Niklas Luhman: a nova teoria dos sistemas. Porto Alegre: UFRGS, 1997. p. 49-59.

MACHADO, A. Tribunais de contas estaduais: indicações, perfil dos conselheiros e autonomia das instituições. Dissertação (Mestrado em Ciência Política) - Universidade Federal de Pernambuco, CFCH, Recife, 2017.

MEDAUAR, O. Controle da administração pública. 3. ed. rev., atual. e ampl. São Paulo: Revista dos Tribunais, 2014.

MIRANDA, A. Tribunal de Contas no Brasil: a falsa cisão entre técnica e política: 1890-2018, da legislação simbólica ao gerencialismo público do ajuste fiscal. Rio de Janeiro: UFRJ, 2020.

NUNES, M. G. Jurimetria: como a Estatística pode reinventar o Direito. 2. ed. rev. e atual. São Paulo: Revista dos Tribunais, 2019.

RECH, C.; ZORTÉA, C. H. C. Domínio contestado: reflexões acerca da corrupção e o papel dos tribunais de contas. In: LIMA, E. C. P. (coord.). *Tribunal de Contas do século XXI*. Belo Horizonte: Fórum, 2020. p. 49-60.

REGATIERI, R. O monitoramento e a avaliação de políticas públicas. In: SALTO, F. S.; PELLEGRINI, J. A. (orgs.). *Contas públicas no Brasil*. São Paulo: Saraiva, 2020. p. 365-394.

RIBEIRO, R. J.; BLIACHERIENE, A. C.; SANTANA, J. L. Considerações sobre formas, processos e isomorfismo nas estruturas de controle interno da federação brasileira. In: BLIACHERIENE, A. C.; BRAGA, M. V.; RIBEIRO, R. J. *Controladoria no setor público*. 2. ed (rev., ampl. e atual.). Belo Horizonte: Fórum, 2019.

ROCHA, D. G. As duas faces de Jano dos Tribunais de Contas Brasileiros no Auxílio ao Exercício do Controle Social sobre a Administração Pública. Tese (Doutorado em Administração Pública e Governo) - FGV-SP, São Paulo, 2017.

RODRIGUEZ, J. R. Como decidem as cortes?: para uma crítica do direito (brasileiro). Rio de Janeiro, FGV, 2013.

ROQUE, A. V.; SANTOS, L. B. R. Inteligência artificial na tomada de decisões judiciais: três premissas básicas. *Revista Eletrônica de Direito Processual*, Rio de Janeiro, v. 22, n. 1, p. 58-78, 2021.

SAKAI, J.; PAIVA, N. Quem são os conselheiros dos Tribunais de Contas? Disponível em: <http://www.transparencia.org.br/downloads/publicacoes/TBrasil%20-%20Tribunais%20de%20Contas%202016.pdf>. Acesso em 4 dez. 2016.

SCAFF, F. F. Orçamento republicano e liberdade igual: ensaio sobre direito financeiro, república e direitos fundamentais. Belo Horizonte: Fórum, 2018.

SCHWAB, K; DAVIS, N. Aplicando a quarta revolução industrial. São Paulo: Edipro, 2019. E-book.

SMULOVITZ, C.; PERUZZOTTI, E. Societal and horizontal controls: two cases of a fruitful relationship. In: MAINWARING, S.; WELNA, C. *Democratic accountability in Latin America*. New York: Oxford University Press, 2003. p. 309-331.

SPECK, B. W. Estratégia para melhorar o relacionamento entre Tribunais de Contas Estaduais com a Sociedade Civil: uma estratégia para aumentar a eficácia da auditoria governamental nos Estados brasileiros. Salvador: Revista do TCE Bahia, 2013. p. 91-99.

SPECK, B. W. Inovação e rotina no Tribunal de Contas da União: o papel da instituição superior de controle financeiro no sistema político-administrativo do Brasil. São Paulo: Fundação Konrad Adenauer, 2000.

TORRES, R. L. Tratado de direito constitucional financeiro e tributário. 3. ed. Rio de Janeiro: Renovar, 2008. v. 5: O orçamento na Constituição.

VILLAVERDE, J.; TEIXEIRA, M. O dilema do controle sobre os controladores. *Revista do Ministério Público de Contas do Paraná (MPC-PR)*, v. 3, p. 02, 2017.

CONTROLE DE SUSTENTABILIDADE (T)EC(N)OLÓGICO PELOS TRIBUNAIS DE CONTAS DO BRASIL E DA ESPANHA E UM BREVE DIÁLOGO COM CERVANTES

Daniela Zago Gonçalves da Cunda.

Pós-doutoranda na Universidade de São Paulo (USP/EACH). Doutora e mestre em Direito pela PUC/RS. Conselheira-Substituta e Presidente da Comissão Permanente de Sustentabilidade do Tribunal de Contas do Estado do Rio Grande do Sul (TCE/RS). Membro do Comitê de Meio Ambiente e Sustentabilidade do Instituto Rui Barbosa.

A sintonia entre os *controles de sustentabilidade externo, interno e social* da Administração Pública deverá ser equivalente à parceria de *Dom Quixote* e *Sancho Pança* abordada por Miguel Cervantes no célebre romance *O Engenhoso Fidalgo Dom Quixote de La Mancha*. Já a sustentabilidade em suas múltiplas dimensões (para além da dimensão fiscal e *com ênfase à dimensão ambiental*) e a utilização das novas tecnologias deveriam ser *Dulcineia de Toboso*, a inspirar e ser idealizada em cada passo dado na longa caminhada rumo a uma maior *equidade intra e intergeracional*, que deverá ser concretizada e não se manter meramente fictícia (nesse ponto diverso do romance em diálogo). No diagnóstico realizado na pesquisa até o presente momento, sucintamente apresentado neste artigo, depreende-se que os controles dispõem de ferramentas tecnológicas para o monitoramento da *dimensão ambiental da sustentabilidade* tanto com a valentia como com fragilidades de *Rocinantes*, devendo, pois, serem aprimorados para que efetivamente tornem-se “fortes alazões”, não obstante a importância que já desempenhem na jornada.

Persistem os debates mundiais pelos países que aderiram ao Acordo de Paris em como estão cumprindo suas metas

estabelecidas de maneira a contornar o aquecimento global, assim como quanto à implementação dos 17 objetivos de desenvolvimento sustentável da Agenda da ONU para 2030. Trata-se de assunção de responsabilidades importantes, mas que não poderão ficar restritas aos pactos assumidos pelos Estados, deverão também ser levadas a sério pelos demais entes federativos, pelas várias instituições públicas e pela própria sociedade.

Os Tribunais de Contas detêm papel relevante nesse cenário de implementação e fiscalização de políticas públicas a promoverem os ODS da Agenda 2030 da ONU e, por consequência, na averiguação do conjunto de *deveres/direitos constitucionais* previstos na Constituição Federal e detalhados em leis infraconstitucionais interligados aos temas previstos com destaque à dimensão ambiental da sustentabilidade. A referida atuação não deverá ser apenas repressiva, mediante um controle que por vezes chega tarde e/ou opressor, mas sim com *deferência* aos gestores probos e eficientes.

Como pressupõe o título, o presente estudo visa a responder a seguinte indagação central: nas possibilidades de atuações dos Tribunais de Contas até que ponto as novas tecnologias e ações inovadoras estão sendo utilizadas para potencializar a implementação, e respectivo controle, de diretrizes de sustentabilidade e do aprimoramento de políticas públicas de longo prazo e para além de fronteiras objetivando o bem-estar da sociedade e do ambiente ecologicamente equilibrado?

No itinerário metodológico, optou-se, quanto aos fins, por pesquisa de natureza descritiva¹ e, quanto aos meios, pela revisão da literatura e pesquisa documental. A investigação segue, portanto, predominantemente a abordagem qualitativa. Como referido, utiliza-se a revisão bibliográfica e documental como método para

¹ GIL, Antônio Carlos. *Métodos e Técnicas de Pesquisa Social*. 7. ed. 2. reimpr. São Paulo: Atlas, 2021.

exame das possibilidades e limites da atuação dos Tribunais de Contas na concretização da *dimensão tecnológica e/ou digital* da sustentabilidade (como instrumento para uma tutela mais ampla das dimensões ecológica e social da sustentabilidade), à luz dos ODS da Agenda da ONU para 2030, da Constituição Federal e de leis infraconstitucionais, como a recente Lei do Governo Digital. O estudo, conjuntamente, envolve pesquisa de campo mediante levantamento junto aos Tribunais de Contas do Brasil (em desenvolvimento) quanto à implementação dos ODS instrumentais n.º 09, 16 e 17 da Agenda 2030 da ONU, seguida de divulgações de boas práticas de ações tecnológicas, inovadoras, dialógicas, pedagógicas e transparentes concedendo resultados positivos, institucionais e sociais, para as gerações presentes e futuras, bem como na concretização de um *Estado Democrático de Direito e também Ecológico*.

No Brasil, a sustentabilidade consta de forma clara como um *dever constitucional* na Constituição Federal, com destaque aos artigos 225 e 170, inc. VI, bem como o artigo 3.º² Vista como dever constitucional ou como *metaprincípio* ou ainda como *garantia constitucional*, torna-se evidente que a *sustentabilidade redefine o papel e as funções do Estado agregando uma missão de curadoria tanto ao Estado como à sociedade*³, também assentada “na lógica de solidariedade responsável inerente ao Estado Social.”⁴

² Nesse sentido, vide também: FREITAS, Juez. Constituição e sustentabilidade: políticas públicas baseadas em evidências e o direito ao bem-estar de presentes e futuras gerações. Interesse Público - IP, Belo Horizonte, ano 24, n. 133, p. 17-27, maio/jun. 2022.

³ BOSSELMANN, Klaus. *O princípio da sustentabilidade: Transformando direito e governança*. Tradução: Phillip Gil França. São Paulo: Revista dos Tribunais, 2015, pp. 23 e ss.

⁴ GOMES, Carla Amado. *Risco e Modificação do Acto Autorizativo Concretizador de deveres de Protecção do Ambiente*. Coimbra: Coimbra Editora, 2007 (versão Ebook), p 151. Também da mesma autora, vide: GOMES, Carla Amado. *Introdução do Direito do Ambiente*, 2.ª ed. AAFDL, 2014. GOMES, Carla Amado. *Sustentabilidade ambiental: missão impossível?* Publicações da FDUL/ICJP, maio/2014. disponível no site:

O *direito/dever constitucional de sustentabilidade*, com destaque às suas dimensões ambiental e ecológica, com atenção às urgências climáticas (nos termos do Projeto de Emenda Constitucional à Constituição da República Federativa do Brasil n.º 37/2021)⁵, requer atuação em rede, aditivada de atuações e controles inovadores, em melhor consonância com as *políticas públicas de longo prazo*. Em estudos anteriores, para além do tripé da sustentabilidade - *ambiental, social e econômico*-, dedicou-se ao desenvolvimento da *dimensão fiscal*, que tem relação forte com a atuação dos Tribunais de Contas⁶. No presente estudo, aborda-se a verificação de como as novas tecnologias poderão auxiliar nessa maior conectividade entre controle externo, social e interno, desenvolvendo-se, pois, a *dimensão tecnológica e/ou digital* da sustentabilidade.

<http://www.icjp.pt/sites/default/files/papers/palmas-sustentabilidade.pdf>
(acesso em 15/07/2022).

⁵ Que muito em breve terá *status* explicitamente de *direito fundamental*, uma vez aprovadas as diretrizes constantes na PEC 37/2021. Para maiores detalhes e acompanhamento dos andamentos da referida alteração constitucional vide: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2304959> (Acesso em 03/11/22)

⁶ CUNDA, Daniela Zago Gonçalves da. *Controle de sustentabilidade pelos Tribunais de Contas*. 2016. Tese (Doutorado em Direito) - Faculdade de Direito, Pontifícia Universidade Católica do Rio Grande do Sul, Rio Grande do Sul, 2016. Vide também: CUNDA, Daniela Zago Gonçalves da. *Controle de sustentabilidade pelos Tribunais de Contas e a necessária ênfase à dimensão ambiental*. In: MIRANDA, Jorge; GOMES, Carla Amado; PENTINAT, Susana Borrás (Coord.). *Diálogo Ambiental, Constitucional e Internacional*. Volume 10, E-Book Internacional. Lisboa: Faculdade de Direito da Universidade de Lisboa (CJP e CIDP), abril de 2020, pp. 293-341. CUNDA, Daniela Zago Gonçalves da. *Controle de Sustentabilidade Fiscal pelos Tribunais de Contas: tutela preventiva da responsabilidade fiscal e a concretização da solidariedade intergeracional*. In: LIMA, Luiz Henrique; SARQUIS, Alexandre (coord.) *Contas Governamentais e Responsabilidade Fiscal: desafios para o controle externo*. Estudos de Ministros e Conselheiros Substitutos dos Tribunais de Contas. Belo Horizonte: Fórum, 2017. CUNDA, Daniela Zago Gonçalves da. *Sustentabilidade Fiscal sob a Ótica da Solidariedade e os Direitos Sociais em Xequê*. *Revista da Faculdade de Direito de Lisboa*, 2013-a, ano 2, n. 3, 1911-1967. Disponível em: <http://www.idb-fdul.com/> ISSN: 2182-7567.

A presente pesquisa pretende aprimorar estudos anteriores que procuraram definir e delimitar o que se entendia por *ciber@ministração pública e controle externo 4.0*.⁷ Na presente proposta, de maneira a tutelar a *dimensão ecológica*, analisam-se as instituições sob uma perspectiva inédita e mais específica, ou seja, a *administração pública e o controle externo tecnológicos* visualizados como *curadores* de um *Estado Democrático de Direito e também Ecológico*, sintetizado na expressão *Estado (t)Ec(n)ológico*, ou seja, um Estado (com políticas públicas de longo prazo) tecnológico e a serviço da ecologia (que engloba os seres humanos, e respectivo bem-estar, como partícipes de um todo maior).

Na evolução do *iter tecnológico* ao longo das últimas décadas, do 1.0 ao 4.0, vários foram os registros de externalidades negativas ao meio ambiente, chegando o momento de as novas tecnologias e ações inovadoras (em um sentido mais abrangente) estarem a serviço da sustentabilidade (multidimensional, com ênfase na dimensão ecológica).⁸ A *Indústria 4.0* ou “fábricas inteligentes”,

⁷CUNDA, Daniela Zago G.; RAMOS, Letícia A. A *Ciber@ministração Pública e Controle 4.0*, seus desafios em tempo de pandemia do coronavírus, e a transparência ampliada (para além de translúcida). In: LIMA, Luiz Henrique et al. (coord.) *Os Desafios do Controle Externo diante da pandemia da COVID 19: Estudos dos Ministros e Conselheiros Substitutos dos Tribunais de Contas*. Belo Horizonte: Fórum, 2020. Tema anteriormente desenvolvido em capítulo específico constante no seguinte estudo: CUNDA, Daniela Zago Gonçalves da. *O Dever Fundamental à Saúde e o Dever Fundamental à Educação na Lupa dos Tribunais (para além) de Contas*. Ebook, Porto Alegre: Editora Simplíssimo Livros, 2013.

⁸ Tal como afirmado no projeto de pesquisa apresentado junto à USP/EACH, com respaldo nos referenciais teóricos debatidos no Grupo de Pesquisas. Para maiores detalhes, vide: CUNDA, Daniela Zago G. da. *A atuação em rede do Controle Externo 4.0 e da Ciber@ministração Pública na concretização do dever constitucional de sustentabilidade e dos ODS da Agenda 2030 da ONU*. Disponível em: <http://www5.each.usp.br/wp-content/uploads/2021/10/POS-DOUTORANDOS-EACH-site-EACH-Fevereiro-2022.pdf> (acesso em 02/11/2022). Com amparo nos seguintes estudos: GONGA, Yiwei; YANGB Jun; SHIA Xiaojie. Towards a comprehensive understanding of digital transformation in government: Analysis of flexibility and enterprise architecture. *Government Information*

com máquinas e insumos que “conversam entre si” ao longo da linha de produção, com verdadeira transformação do modo de produzir e de consumir, com dados e informações com tratamento prioritário e decisivo, são consideradas como antecedentes lógicos das *Smart Cities*⁹.

Neste ponto, cabível uma indagação central: As ferramentas tecnológicas e ações inovadoras (em um sentido mais amplo) utilizadas pelos Tribunais de Contas têm por objeto a fiscalização e tutela da dimensão ecológica da sustentabilidade ou destinam-se com primazia à averiguação de direitos fundamentais interligados à dimensão social e fiscal da sustentabilidade? A seguir, serão sucintamente enfrentados tais questionamentos.

Apresentar-se-á uma síntese dos resultados no atual estado da arte de pesquisas de campo realizadas junto aos 33 Tribunais de Contas do Brasil, com panorama inicial quanto à implementação dos ODS instrumentais n.º 09, 16 e 17 da Agenda 2030 da ONU, seguido de divulgações de boas práticas de ações tecnológicas e inovadoras.

Quarterly, n. 37, 2020, 101487, pp. 1 – 13, disponível no seguinte site: <https://doi.org/10.1016/j.giq.2020.10148> (acesso em 12/09/2022); PITTAWAYA, Jeffrey J.; MONTAZEMIB, Ali Reza. Know-how to lead digital transformation: The case of local governments. *Government Information Quarterly*. Volume 37, Issue 4, October 2020, disponível no seguinte site: <https://doi.org/10.1016/j.giq.2020.101474> (acesso em 12/09/2022); POSTER, Mark. *CyberDemocracy: Internet and the Public Sphere*. Disponível em: www.forumglobal.de/soc/bibliot/p/cyberdemocracy.poster.htm. Acesso em 09/07/2022; PUTRIJANTI, Aju. Participation of Society In Decision-Making By Government In Industrial Revolution 4.0. *E3S Web of Conferences* 125, (2019), disponível no seguinte site: <https://doi.org/10.1051/e3sconf/20191250> (acesso em 12/09/2022)

⁹ Sobre o tema, vide: MENDES, Alexandre J; ROSA, A. M.; SOUZA, Maria Claudia da Silva Antunes de. Sociedade 5.0, Smart Cities e a Indústria 4.0. *In Direito e sustentabilidade I*. TYBUSCH, Jerônimo Siqueira; SOARES, Josemar Sidinei; SOUZA, Maria Claudia da Silva Antunes De (coord.) – Florianópolis: CONPEDI, 2021, pp. 236-256.

Pesquisa finalizada em 2021¹⁰, tendo como ponto de partida levantamento realizado junto aos 33 Tribunais de Contas do Brasil, com a ferramenta *LimeSurvey*, da análise das respostas obtidas de 13 Tribunais de Contas¹¹, correspondendo a 39,39%¹², constatou-se

¹⁰ COMISSÃO DE SUSTENTABILIDADE DO TCE/RS; CUNDA, Daniela Z. G. (Presidente da Comissão TCE/RS e Relatora das Pesquisas); RAMOS, Leticia (Coordenadora do Grupo Temático do Diagnóstico/Pesquisa e Relatora); FRANSCSCHINI, Carina L. (Relatora das Pesquisas). *Relatório da Pesquisa de adesão da Agenda 2030 pelos Tribunais de Contas*. Comissão de Sustentabilidade, 2021.

<https://portalnovo.tce.rs.gov.br/temas-especiais/resultado-pesquisa-adesao-agenda-2030/>

A pesquisa realizada foi uma das ações propostas junto à Comissão Permanente de Sustentabilidade com objetivo de verificar a aderência dos Tribunais de Contas, Ministério Público de Contas e Instituições acerca dos objetivos sustentáveis ou áreas temáticas correlatas ao tema, bem como verificar os principais instrumentos utilizados para a implementação dos temas/ODS. Além disso, o levantamento permitiu identificar espaços para fortalecimento de parcerias para implementação dos objetivos, medida que vai ao encontro do disposto no ODS 17 e, por consequência, dos demais.

¹¹ 13 Tribunais de Contas Estaduais (TCEs), correspondendo a 39,39%: TCE-AC, TCE-AP, TCE-AM, TCE-CE, TCE-ES, TCE-MG, TCE-PA, TCE-PB, TCE-RJ, TCE-RS, TCE-RO, TCE-SC, TCE-SP.

¹² Os resultados constantes no relatório referido nas notas anteriores podem ser sistematizados em três principais eixos. O *primeiro eixo*, relacionado à *governança sustentável*, propôs-se um levantamento das ações dos Tribunais de Contas em seu âmbito interno, ou seja, no sentido de averiguação dos Tribunais de Contas como gestores comprometidos com a efetivação dos ODS, acrescidos de um corpo técnico capaz de desenvolvê-las.

No que tange ao primeiro eixo de atuação, *referente à existência de iniciativas vinculadas aos ODS no ambiente interno*, dos 13 respondentes, 11 informaram que sim, correspondendo a 84,62%. Nesse âmbito, cumpre ser ressaltado que os dois que responderam não, correspondendo a 15,38%, quando questionados quanto aos temas envolvidos nos ODS (v.g. saúde, educação, licitações sustentáveis) um dos Tribunais detalhou iniciativas adotadas. Em um *segundo eixo*, dando destaque ao papel pedagógico e dialógico a fomentar a concretização dos ODS junto a outras instituições e a sociedade, dos 13 Tribunais de Contas que responderam, 9 (nove) Cortes de Contas sinalizaram que sim. Novamente é importante a ressalva de que 4 (quatro) dos Tribunais de Contas que responderam não possuir atuação, em parte, detectou-se atuações pedagógicas, por exemplo. Quanto ao *terceiro eixo de atuação, referente aos jurisdicionados fiscalizados*, no que tange ao questionamento “há iniciativas do seu TC para promover os ODS?”, dos 13 (treze)

em comparativo levando em consideração os 17 ODS da Agenda da ONU para 2030, *averiguando-se os ODS instrumentais*, o seguinte ranking de prioridades na implementação dos ODS *no âmbito interno dos Tribunais de Contas*, contando com ações já implementadas, os ODS 16, 17 e 09 ocupam respectivamente as 5ª, 7ª e 9ª posições. Na implementação (em andamento na época de testagem) *junto aos jurisdicionados*, os ODS 16, 17 e 09 figuraram com as seguintes e respectivas prioridades 5ª, 6ª e 8ª colocações. Já na averiguação de sistemáticas interligadas a inovações e conectividade *junto à sociedade*, às ações em andamento para adesão ou apenas com previsão de implementação dos ODS 16, 17 e 09, encontraram-se em um preocupante menor grau de prioridade (tendo em mente os desejáveis *controles externo e social 5.0*), ou seja, com as respectivas colocações: 6ª, 16ª e 15ª posições).¹³

No transcorrer do segundo semestre de 2022 iniciou-se uma segunda pesquisa de campo, junto aos setores de informática dos 33 Tribunais de Contas do Brasil. Considerando as informações com especificidades tecnológicas almejadas, inicialmente foram encaminhados questionários diretamente para servidores especializados em tecnologia da informação, não se obtendo, todavia, os retornos esperados, em razão das limitações (v.g. setores especializados para atendimento de demandas internas) e ausência de autorização institucional para que as respostas fossem fornecidas. Realizou-se, na sequência, uma terceira etapa da

Tribunais de Contas que responderam o questionário, 10 (dez) mencionaram ter iniciativas. Dos 3 (três) que responderam ainda não terem atuação específica nesse âmbito, ao serem questionados quanto às áreas temáticas interligadas aos ODS, 2 (dois) responderam que sim, apenas um dos Tribunais de Contas referiu não possuir atuação junto aos jurisdicionados.

¹³ COMISSÃO DE SUSTENTABILIDADE DO TCE/RS; CUNDA, Daniela Z. G. (Presidente da Comissão TCE/RS e Relatora das Pesquisas); RAMOS, Letícia (Coordenadora do Grupo Temático do Diagnóstico/Pesquisa e Relatora); FRANSCSCHINI, Carina L. (Relatora das Pesquisas). *Relatório da Pesquisa de adesão da Agenda 2030 pelos Tribunais de Contas*. Comissão de Sustentabilidade, 2021. <https://portalnovo.tce.rs.gov.br/temas-especiais/resultado-pesquisa-adesao-agenda-2030/>

pesquisa, junto às Ouvidorias dos Tribunais de Contas do Brasil, viabilizando-se, conjuntamente, uma averiguação do pronto atendimento e cumprimento da Lei de Acesso a Informações pelos próprios Tribunais de Contas, instituições que fiscalizam a transparência de outras instituições públicas. Na pesquisa em desenvolvimento, junto aos 33 Tribunais de Contas do Brasil, foram encaminhados os seguintes questionamentos:

1) O Tribunal de Contas utiliza novas tecnologias e/ou ações inovadoras nas suas atuações institucionais (como controle de políticas públicas em geral, no monitoramento de receitas/despesas públicas e/ou licitações e contratos)? 1.1) Se sim, quais as ações inovadoras implementadas e quais as ferramentas tecnológicas utilizadas (mediante sucinto relato)? 1.2) Melhor detalhando, o Tribunal de Contas utiliza Inteligência Artificial, Internet das Coisas - IoT, Big Data, Blockchain, Machine Learning, Drones, dentre outras ferramentas tecnológicas no controle de políticas públicas ou demais atividades institucionais?

2) O Tribunal de Contas utiliza novas tecnologias e/ou ações inovadoras no controle de políticas públicas relacionadas à proteção do meio ambiente (como por exemplo, monitoramento de licenciamentos, biomas, recursos hídricos, resíduos sólidos, saneamento básico, poluição, licitações sustentáveis, eficiência energética e/ou mudanças climáticas)? 2.1. Se sim, quais as ações inovadoras implementadas e quais as ferramentas tecnológicas utilizadas (mediante sucinto relato)? 2.2) Melhor detalhando, o Tribunal de Contas utiliza Inteligência Artificial, Internet das Coisas - IoT, Big Data, Blockchain, Machine learning, Drones, dentre outras ferramentas tecnológicas no controle de políticas públicas relacionadas à proteção do meio ambiente?

3) **A Ouvidoria de seu Tribunal de Contas recebe denúncias relacionadas à proteção do meio ambiente** (como por exemplo, monitoramento de licenciamentos, biomas, recursos hídricos, resíduos sólidos, saneamento básico, poluição, licitações sustentáveis, eficiência energética e/ou mudanças climáticas)? 3.1. Se sim, seria possível (mesmo que por estimativa) mencionar o percentual dessas denúncias? 3.2. Quais os encaminhamentos realizados (mediante sucinto relato)? 3.3) Melhor detalhando, **o Tribunal de Contas na ações institucionais (interligadas à Ouvidoria) utiliza Inteligência Artificial, Internet das Coisas - IoT, Big Data, Blockchain, Machine learning, Drones, dentre outras ferramentas tecnológicas?**

Não obstante a terceira etapa da pesquisa esteja em desenvolvimento, há possibilidade de serem divulgadas algumas boas práticas com destaque a algumas ferramentas tecnológicas e ações inovadoras e que potencializam conexões. Dos 33 Tribunais de Contas sondados, no prazo estabelecido na Lei de Acesso à Informação, 21 deram retorno, ou informando de imediato às respostas aos questionamentos enviados¹⁴, ou motivando a necessidade de maior prazo para atenderem a demanda (v.g. levantamento manual para responder o item 3, dentre outras justificativas interligadas ao fato de as respostas envolverem múltiplos setores da instituição).¹⁵

Quanto ao primeiro questionamento - *o Tribunal de Contas utiliza novas tecnologias e/ou ações inovadoras nas suas*

¹⁴ Com merecido destaque ao pronto atendimento prestado aos cidadãos pela Ouvidoria dos seguintes Tribunais de Contas: TCE/CE, TCE/PA, TCE/AP e TCM/GO.

¹⁵ Quanto aos Tribunais de Contas que não responderam prontamente ao questionário enviado para a Ouvidoria, muito possivelmente em razão de seus volumes de demandas, serão providenciados contatos alternativos (v.g. mediante solicitação de audiências e reuniões) assim como ampliação de prazo com o intuito de obter valiosas contribuições complementares.

atuações institucionais – dos 4 respondentes, 3 afirmaram utilizar ferramentas tecnológicas, tendo sido informado por um dos Tribunais de Contas que **são utilizados modelos baseados em Machine Learning para predição de receitas a partir de um Big Data desenhado para tal finalidade**

Referente ao segundo questionamento - o *Tribunal de Contas utiliza novas tecnologias e/ou ações inovadoras no controle de políticas públicas relacionadas à proteção do meio ambiente* - nenhum dos Tribunais de Contas do Brasil que deu retorno até o presente momento utiliza ações inovadoras ou ferramentas tecnológicas no controle de políticas públicas interligadas à dimensão ecológica e ambiental da sustentabilidade. Saliente-se que um dos Tribunais de Contas indicou haver convênio de desenvolvimento de IA com Universidade Federal.

E por fim, quanto às indagações - a Ouvidoria de seu Tribunal de Contas recebe denúncias relacionadas à proteção do meio ambiente - a resposta foi negativa por todos os respondentes até o presente momento, sendo que indicaram ter procedimentos e tramitações adequadas para atender eventuais demandas.¹⁶

Pertinente ao questionamento - O Tribunal de Contas nas ações institucionais (interligadas à Ouvidoria) utiliza ferramentas tecnológicas e/ou ações inovadoras? - a totalidade das instituições que já concluíram a participação na presente pesquisa afirmou não utilizar ferramentas tecnológicas como as exemplificativamente indicadas. Por outro lado, foi informado por um dos entrevistados que

¹⁶ Como exemplificativamente a Resolução n.º 03/2022 e Portaria de funcionamento da Ouvidoria do TCE/CE. Para maiores detalhes, vide: <https://www.tce.ce.gov.br/exercicios-antiores/resolucoes-administrativas> (Acesso em 06/11/2022).

atualmente trabalham com sistemas de informações aprimorando constantemente os mesmos com o objetivo de inovação. Já outro respondente indicou haver convênio de desenvolvimento de IA com Universidade Federal.¹⁷

Depreende-se na averiguação mais detalhada e investigativa junto aos Tribunais de Contas do Brasil, que no atual “estado da arte” as novas tecnologias estão prioritariamente direcionadas ao controle do gasto (dimensão fiscal) e a reduzir a corrupção; menores são as utilizações das ferramentas tecnológicas diretamente para tutelar a dimensão social e ainda mais tímidas são as inovações para salvaguardar a dimensão ecológica da sustentabilidade, nesse âmbito merecendo destaque a boa prática do TCE/AM o “APP sou Eco”.¹⁸

Diante de dados disponíveis nos sites dos Tribunais de Contas pesquisados, Instituto Rui Barbosa e ATRICON, assim como das respostas passíveis de divulgação até o presente momento da pesquisa em desenvolvimento, as seguintes boas práticas merecem ser destacadas a título exemplificativo:

Convênio da ATRICON e o MapBiomias, para utilização dos dados e plataformas MapBiomias fortalecendo a transparência sobre as ações dos órgãos de controle e combate ao desmatamento, além de aprimorar a eficácia

¹⁷ Conforme relatório de pesquisa detalhado: CUNDA, Daniela Zago Gonçalves da. *Relatório do Diagnóstico sobre utilização de novas tecnologias e inovações nas ações institucionais dos Tribunais de Contas do Brasil e respectivas Ouvidorias*, desenvolvido junto ao Grupo de Pesquisas SmartCitiesBR na Universidade de São Paulo (USP/EACH), 2022 e 2023.

¹⁸ Para maiores detalhes sobre a ferramenta, vide: https://play.google.com/store/apps/details?id=com.ecocidadao&hl=pt_BR&gl=US (Acesso em 06/11/2022). Já a outra página do aplicativo indicada no seguinte site: <https://soueco.tce.am.gov.br> (encontrava-se indisponível).

dos sistemas da administração pública no enfrentamento dos ilícitos ambientais;¹⁹

Aplicativo do TCE/AM Sou Eco - O aplicativo "Sou Eco" é uma tecnologia que visa oferecer à população um canal fácil e rápido para fazer denúncias de infrações e crimes ambientais, colaborando com o trabalho dos órgãos fiscalizadores. O aplicativo permite realizar denúncia anônima e denúncia cadastrada. As denúncias serão enviadas para os órgãos fiscalizadores mais próximos de acordo com o tipo de denúncia. O aplicativo está disponível nos idiomas: Inglês, Espanhol, Francês e Português;²⁰

Monitor Fiscal do TCE/CE, que contempla a criação de painéis interativos para o monitoramento das seguintes áreas: receitas, despesas, dívida e previdência;²¹

Monitor Fiscal Nacional do IRB, nos moldes das tecnologias desenvolvidas pelo TCE/CE disponibilizadas às demais instituições e cidadãos;

O TCM/GO vem desenvolvendo a **ferramenta COLARE/MESTRA**, que trouxe novas tecnologias e evoluções tecnológicas no controle das licitações e

¹⁹ <https://atrimon.org.br/tribunais-de-contas-firmam-acordo-para-ajudar-na-conservacao-do-meio-ambiente/> (Acesso em 23/04/23)

²⁰ Sobre a ferramenta, vide:

https://play.google.com/store/apps/details?id=com.ecocidadao&hl=pt_BR&gl=US (Acesso em 06/11/2022) Já a outra página do aplicativo indicada no seguinte site: <https://soueco.tce.am.gov.br> (encontrava-se indisponível).

²¹ Para maiores informações da plataforma indica-se consultar o endereço eletrônico <https://monitorfiscal.tce.ce.gov.br/> (acesso em 06/11/2022).

contratos, bem como dos atos de pessoal e folha de pagamento;²²

Laboratório de Informações de Controle (LabContas)²³;

Rede Nacional de Informações Estratégicas (InfoContas)²⁴;

Diretrizes traçadas pela Associação dos Tribunais de Contas do Brasil - ATRICON, na Resolução nº 09/2018, de maneira a sistematizar a fiscalização e demais coordenadas

²² O COLARE conta com o robô ESMERALDA, que realiza a auditoria eletrônica de todas as licitações e contratos enviados ao Tribunal, hoje já temos quase 3.000.000 de envios, que são auditados por um conjunto de regras e pontos de controles pré-definidos (mais de 80), que ao final da análise recebem uma classificação de riscos que irá auxiliar o auditor na definição de qual achado será relevante realizar uma auditoria mais aprofundada. Mais informações sobre a plataforma COLARE podem ser obtidas em <https://www.tcmgo.tc.br/enviodedados/> (Acesso em 06/11/2022).

²³ Plataforma com centena de bases de dados resultantes de acordos de cooperação entre várias instituições públicas e o TCU, com informações que têm como destinatários os partícipes do controle externo, interno e social. Para maiores detalhes, vide a Portaria n.º 102/2020 do TCU, com destaque ao art. 4.º que elenca os objetivos, como o de garantir informações destinadas ao cumprimento da missão institucional dos Tribunais de Contas (para além de quantidade, também visando a qualidade e tempestividade). A lógica de fomentar o intercâmbio de informações e proporcionar a exploração de dados, mediante uso de recursos de tecnologia também se encontra presente, nos termos recomendados na Lei do Governo Digital. <https://portal.tcu.gov.br/lumis/portal/file/fileDownload.jsp?fileId=8A81881E72E6BE55017305B43F4A3157&inline=1> Acesso em 18/02/2023.

²⁴ Estabelecida mediante Acordo de Cooperação Técnica entre a ATRICON (Associação Nacional dos membros dos Tribunais de Contas do Brasil), o Instituto Rui Barbosa (IRB), TCU, TCE's e TCM's, com diretrizes aprovadas pela Resolução n.º 07/2014. Iniciativa que também poderá contribuir para o aperfeiçoamento dos Portais Eletrônicos de Transparência Fiscal, assim como na implementação das diretrizes constantes na Lei do Governo Eletrônico. Para maiores detalhes, vide: <https://irbcontas.org.br/infocontas/> Acesso em 18/02/2023.

quanto à Transparência dos Tribunais de Contas e dos jurisdicionados;

Ferramentas tecnológicas do Tribunal de Contas do Rio Grande do Sul sobre transparência ativa dos municípios, que demonstra a evolução e a melhora dos repositórios oficiais de informação²⁵

Levantamento²⁶ realizado pela entidade Transparência Internacional que avaliou o nível de disponibilização das informações sobre contratações emergenciais dos entes da federação, sobre a temática referente ao combate à Covid-19.

No levantamento realizado no Tribunal de Cuentas²⁷, ainda em estágio inicial e em conformidade com os dados disponíveis a todos os cidadãos em seu sítio eletrônico²⁸, assim como no Planejamento Estratégico²⁹, constata-se junto aos principais objetivos: fomentar boas práticas de organização, gestão e controle das entidades públicas; identificar e fiscalizar as principais áreas de risco; incrementar as fiscalizações operativas; intensificar as relações de coordenação, cooperação e colaboração com entidades fiscalizadoras superiores e suas organizações

²⁵ http://portal.tce.rs.gov.br/docs/transparencia_2018/relatorio_transparencia_2018.pdf. Acesso em 18/02/2023.

²⁶ <https://irbcontas.org.br/ranking-indica-bons-indices-de-transparencia-nos-gastos-com-a-covid-19-em-estados-e-municipios/>. Acesso em 18/02/2023.

²⁷ <https://www.tcu.es/es> Acesso em 23/04/2023. "Su regulación se encuentra contenida en La Constitución Española, en La Ley Orgánica del Tribunal de Cuentas de 1982 y en La Ley de Funcionamiento del Tribunal de Cuentas de 1988, además de en otras leyes específicas. Es un órgano que depende directamente de las Cortes Generales y ejerce sus funciones por delegación de las mismas en el examen y comprobación de la Cuenta General del Estado, Y cuyo presupuesto se integra em lós Presupuestos Generales Del Estado."

²⁸ <https://www.tcu.es/es/la-institucion/Documentos-Oficiales> Acesso em 23/04/2023.

²⁹ https://www.tcu.es/export/sites/default/galleries/pdf/PLAN ESTRATEGIC O_2018-2021.pdf Acesso em 23/04/2023.

internacionais; reforçar a cooperação e a coordenação do sistema de controle externo na Espanha; contribuir para o controle interno; intensificar as relações com âmbitos institucionais, acadêmicos e profissionais; promover a comunicação interna e externa da atividade de Tribunais de Contas; *facilitar o acesso à informação da instituição*³⁰; implantar um sistema global de garantia da qualidade; reforçar a gestão dos recursos humanos e materiais do Tribunal de Cuentas com atenção ao princípio da eficiência, economia, igualdade³¹ e transparência, tendo como objetivo específico (4.4) modernizar os procedimentos internos e impulsionar os projetos de administração eletrônica e a utilização das *Tecnologias de Informação* para reforçar a eficácia e eficiência do Tribunal de Contas³².

³⁰ M 1: Reorganizar y sistematizar el contenido del sitio web del Tribunal, de forma que facilite el acceso a la información y mejore la divulgación de la actividad institucional, incluyendo, además, de la información exigible por la ley, aquella que pueda ser pública y resulte de interés general M 2: Reorganizar y sistematizar el Portal de Transparencia para simplificar el acceso de los ciudadanos a la información M 3: Incluir en la memoria anual del Tribunal un apartado relativo al grado de cumplimiento de las obligaciones en materia de transparencia M 4: Facilitar la comprensión de los Informes de Fiscalización empleando un lenguaje claro y accesible, e incluyendo un resumen de los objetivos, hallazgos y recomendaciones, que facilite su difusión pública M 5: Dotar a las resoluciones jurisdiccionales de una extensión acorde a la cantidad y complejidad de las cuestiones examinadas y resueltas en las mismas, empleando un lenguaje que resulte claro y accesible.

³¹ Com ênfase na igualdade de gênero (objetivo específico 4.2, M.2). Vide também: <https://www.tcu.es/export/sites/default/galleries/pdf/PLAN-IGUALDAD-Aprob-Pleno-29-09-2014.pdf> (acesso em 23/04/2023)

³² OBJETIVO ESPECÍFICO 4.4: Modernizar los procedimientos internos e impulsar los proyectos de administración electrónica y la utilización de las TIC para reforzar la eficacia y eficiencia en el Tribunal de Cuentas M 1: Elaborar un plan plurianual en el que se determinen las prioridades en el ámbito de las TIC y que comprenda una metodología de selección de proyectos a desarrollar y los procedimientos de gestión de los mismos M 2: Potenciar el uso de herramientas tecnológicas y de procedimientos telemáticos en todos los ámbitos de actuación del Tribunal de Cuentas M 3: Impulsar el desarrollo de una única aplicación informática en el Tribunal de Cuentas que permita un examen más dinámico, homogéneo y efectivo de la Cuenta General del Estado M 4: Elaborar un

Em suas atuações, o Tribunal de Cuentas também se destaca pelo *controle de sustentabilidade fiscal* com atuações menos frequentes no que tange à *dimensão ambiental da sustentabilidade*. Mais recentemente as Instituições Fiscalizadoras revisaram medidas de atuação para reforçar a eficiência do controle e da gestão econômico-financeira pública, com fiscalizações receptivas às demandas da sociedade, dos riscos emergentes, com melhoras na gestão dos fundos públicos, com maior transparência, igualdade, *sustentabilidade ambiental e preparação e implementação dos ODS* da Agenda da ONU para 2030, com destaque aos objetivos interligados às fiscalizações realizadas pelos Tribunais de Contas. Conjuntamente, têm sido crescentes as preocupações com melhoras continuadas das ferramentas eletrônicas utilizadas nos procedimentos fiscalizadores, como forma de possibilitar um controle mais eficiente e eficaz, bem como ampliação de *fiscalizações coordenadas*³³.

O Pleno do Tribunal de Cuentas no Programa de Fiscalização para 2023 aprovou importantes atuações relacionadas aos temas de destaque, v.g.: fiscalização dos *Fondos Next Generation*, planos de

programa de cumplimiento del Esquema Nacional de Seguridad y de la Ley Orgánica de Protección de Datos, em el que se definan las medidas a adoptar, lós hitos y el calendario de su implatación.

³³ Conforme dados constantes no seguinte site: <https://www.tcu.es/es/sala-de-prensa/noticias/Las-Instituciones-fiscalizadoras-revisan-y-proponen-medidas-de-actuacion-para-reforzar-la-eficiencia-del-control-de-la-gestion-economico-financiera-publica/> (Acesso em 23/04/2023). Informam, conjuntamente que “insistieron en la conveniencia de promover la interconexión de sus propias aplicaciones y plataformas electrónicas con otras ya existentes en su entorno de actuación, entre otros, en el ámbito de la contratación pública, al objeto de maximizar la eficiencia en la obtención de la información sobre la gestión y en su control. Igualmente fueron objeto de análisis otros ámbitos en los que estima podrían reforzarse la cooperación y la colaboración institucional; entre otras áreas, en materia de **formación del personal**, así como en el **intercambio de experiencias y buenas prácticas de fiscalización**. De esta forma se contribuirá a potenciar **una red de conocimiento** entre el personal de las Instituciones de control y a homogeneizar criterios, procedimientos y herramientas para el ejercicio de la fiscalización.”

atuações concretas como a prevenção integral da violência de gênero (ODS n.º 05 da Agenda 2030 da ONU), transformação digital e modernização das Entidades Locais que receberam subvenções, avaliação da eficiência energética dos edifícios públicos (ODS n.º 07, 09 e 12 da Agenda 2030 da ONU), destaque especial aos objetivos transversais de igualdade, transparência e sustentabilidade ambiental (que permanece previsto no Plano Estratégico do Tribunal de Contas em duas atividades fiscalizadoras), assim como atuações fiscalizadoras em âmbito internacional como a parceria com Tribunal de Contas de Portugal sobre atuações de defesa contra a desertificação e prevenção e extinção de incêndios florestais na península ibérica (ODS n.º 13, 15, 16 e 17 da Agenda 2030 da ONU).³⁴

Os resultados das *boas práticas* referidas exemplificativamente revelam a importância que os Tribunais de Contas adquirem no sentido de *avaliar e fomentar* o aprimoramento dos *portais de transparência* em consonância com o papel de colaboração e orientação.

Em consonância com estudos que afirmam que a transformação digital é um imperativo estratégico para os governos e instituições que visam à maior eficiência; na mesma linha são as recentes diretrizes legais constantes no ordenamento jurídico, como a Lei n.º 14.129/2021, e as metas internacionais, como as previstas nos ODS (com carácter instrumental) constantes na Agenda 2030 da ONU (ODS n.º 16 - instituições eficazes- e ODS n.º 17 - parceria global-, além do mencionado fomento à inovação constante no ODS 09).

³⁴ Conforme dados constantes no seguinte site: <https://www.tcu.es/es/sala-de-prensa/noticias/El-Pleno-aprueba-el-Programa-de-Fiscalizaciones-para-el-2023/> Acesso em 23/04/2023.

Levando em consideração que a noção de *Sociedade 5.0* centraliza-se na qualidade de vida dos humanos em sociedade³⁵, e que para tal é essencial um ambiente ecologicamente equilibrado, o *controle externo* realizado pelos Tribunais de Contas e o *controle social*, tutelado por essas instituições, deverão ir além de controles 4.0, modernizando-se, como forma de atender as crescentes demandas atinentes às políticas públicas transterritoriais e transtemporais mediante um cibernético e humanitário controle de sustentabilidade 5.0.

Além da fiscalização do cumprimento dos ditames legais de transparência, papel que tem sido desempenhado ao longo das últimas décadas, a atuação das Cortes de Contas deverá ser mais abrangente, averiguando resultados e a qualidade das informações prestadas (v.g. se de fato são informações palatáveis ao cidadão médio), assim como desempenhando o papel de indutor e de provedor da cidadania. Ademais, os portais eletrônicos de transparência dos Tribunais de Contas deverão espelhar a mais eficaz transparência e propiciar que a sociedade tenha condições de fiscalizar o fiscalizador, além de consistirem em verdadeiros bancos de dados para a cidadania (a incluir a comunidade científica).

O presente estudo em desenvolvimento procurou ressaltar algumas boas práticas em matéria de sustentabilidade e novas tecnologias, assim como demonstrar que ainda há um longo caminho a ser percorrido pela Administração Pública e respectivo controle externo, interno e social, na lógica de que o “tecno” esteja

³⁵ Tema em desenvolvimento e a ser *publicado* em estudo mais amplo no transcorrer de 2023, com detalhamento dos resultados posteriormente obtidos em pesquisa de campo realizada junto aos 33 Tribunais de Contas do Brasil em coautoria com a Professora Doutora e supervisora do pós-doutorado junto à USP/EACH - Professora Ana Carla Bliacheriene, com o seguinte título: *Tribunais de Contas (T)Ec(n)ológicos e o Controle de Sustentabilidade 5.0 para uma maior eficiência e equidade de políticas públicas de longo prazo e transterritoriais.*

efetivamente a serviço do ecológico, considerando-se as urgências climáticas.

As diretrizes de sustentabilidade, que incluem a dimensão tecnológica, deverão ser ideais a serem perseguidos no dia a dia da administração, todavia com a cautela de que não sejam apenas fictícios como a figura de *Dulcineia del Toboso*. Embora no início do texto, em diálogo com Cervantes, tenha se destacado a importância da sintonia entre os controles, utilizando-se as figuras de *Dom Quixote* e *Sancho Pança*, os gestores públicos e respectivos controles não poderão confundir utopia e realidade. Dito de outra forma, há que se ter cautela e empenho na efetiva e urgente implementação das dimensões da sustentabilidade e dos Objetivos de Desenvolvimento Sustentável da Agenda da ONU para 2030, com destaque aos interligados às dimensões ambiental e tecnológica da sustentabilidade, sob pena de os Tribunais de Contas exercerem meros simulacros de controle de sustentabilidade (uma espécie de *Greenwashing* ³⁶Control).

³⁶ Nos termos *Greenswashing* abordado no seguinte estudo: BRAGA JÚNIOR, Sergio; MARTINEZ, Marta; CORREA, Caroline; MOURA-LEITE, Rosamaria; SILVA, Dirceu. *Greenwashing effect, attitudes, and beliefs in Green consumption*. 2023, disponível no seguinte site: <https://www.scielo.br/j/rmj/a/j8KWHs8k4XfndmpPCcG9f6f/?lang=en> (Acesso em 23/04/23).

INTELIGENCIA ARTIFICIAL, ALGORITMOS Y TRABAJO

Juan Raso Delgue

Miembro de Número de la Academia Iberoamericana de Derecho del Trabajo y de la Seguridad Social.

INTRODUCCIÓN

Las transformaciones de las relaciones laborales experimentan una nueva vuelta de tuerca, al comprobar que la máquina y la elaboración digital de datos intervienen en la contratación del trabajo, para sustituir decisiones propias del empleador o aportar reglas y criterios nuevos en la organización del trabajo. Ante un tema complejo, la investigación debe focalizar todos sus aspectos - positivos y negativos - porque como expresa Morin, “la realidad es compleja; los fenómenos son complejos y no se pueden explicar a través de afirmaciones radicales, las que pueden conducir a visiones parcializadas de la realidad”¹.

Observamos además la ambivalencia de todas las tecnologías del pasado y del presente. Como el radio descubierto por los esposos Curie, que cura enfermedades y ayudó a construir armas de destrucción masiva, o como la dinamita de Nobel, que abrió galerías y destruyó vidas humanas, los algoritmos informáticos prestan ayuda y crean problemas a la humanidad. A la pregunta de si es bueno o malo que un algoritmo contrate o despida a un trabajador, nuestra sensibilidad humana expresa contrariedad, porque rechazamos la idea que una máquina pueda decidir sobre el vínculo laboral de un ser humano. Pero si preguntáramos a nuestros jóvenes colegas de Facultad si prefieren un tribunal de concurso conformado por docentes (con sus inevitables emociones, preferencias o prejuicios) o bases de dato

¹ Morin E., *Introducción al pensamiento complejo*, Barcelona 1994, p. 34.

tecnológicamente objetivas, probablemente algunos o muchos optarían por decisiones formuladas por algoritmos, cuya imparcialidad parecería estar por encima de las preferencias humanas.

Estas premisas nos preparan para entender e intentar explicar un tema nuevo, que puede aparejar desconcierto, pero que también puede aportar beneficios al ser humano, evitando en todo caso afirmaciones categóricas, que respondan a una idea o relato únicos.

SIGNIFICADO DE ALGORITMO

La expresión “algoritmo” es definida por la Real Academia Española como un “conjunto ordenado y finito de operaciones que permite hallar la solución de un problema”, es decir que es una “secuencia de operaciones”, que nos da como resultado un producto o la solución a un problema planteado.

Hay distintas opiniones sobre el origen de la palabra, de indudable origen árabe como indica el prefijo “al”. La más conocida nos remite a una localidad cerca de Bagdad - Al-Khwārizmī (latinizado luego como Al-Juarismi), donde nació y vivió entre el 780 y el 840 d.C. el célebre matemático Ben Musa, apodado Al-Khwārizmī. Entre sus aportes a Occidente se registra la difusión de los números indo-arábicos. Otra versión indica que la palabra proviene del latín tardío alborarismus, y este a su vez es una abreviación del árabe clásico *ḥisābu lġubār*, que significa ‘cálculo mediante cifras árabigas’. De todos modos, las versiones coinciden sobre el vínculo de la expresión con la matemática musulmana.

La noción de algoritmo tiene aspectos de la teoría sistémica en la medida que tenemos una proposición o datos iniciales (in-put), un proceso de transformación que se traduce en una secuencia de pasos ordenados y un resultado final (out-put). A diferencia de un

“sistema”, el algoritmo se agota en el producto final; en el sistema el producto retroalimenta a los insumos.

Ejemplos comunes de algoritmos en matemática son una multiplicación o una división, el cálculo del mínimo común divisor o reglas para la solución de ecuaciones. En nuestro hábito de aterrizar conceptos abstractos a la vida cotidiana, expresamos que una receta de cocina “es” ... un algoritmo.

En la receta tenemos determinados ingredientes (azúcar, harina, levadura, huevos, etc.), batimos todos ellos con la intención o finalidad de preparar una torta de crema. Pero también es cierto que si modificamos los ingredientes (por ejemplo, agregamos cocoa), tendremos una torta de chocolate. La sencillez del ejemplo nos permite de todos modos entender que el producto final estará siempre vinculado a los ingredientes de la receta. Ello - trasladado al tema que nos ocupará en las próximas páginas - significa que cualquier aplicación de algoritmos al mundo del trabajo podrá ser más o menos acertada, más justa o más discriminatoria, en virtud de los datos (*ingredientes*, en el símil de la torta) que pretendamos utilizar para alcanzar determinado producto. Ello también determina un segundo aspecto vinculado a cualquier estudio sobre los algoritmos y el trabajo: el conocimiento de los datos que constituyen el punto de partida del algoritmo es central para definir la legitimidad del mismo. De ahí que uno de los aspectos más reclamados es el derecho de la persona cuyos datos se usen, a conocer los procedimientos de elaboración del algoritmo, cuestión no siempre visible.

INTELIGENCIA ARTIFICIAL, ALGORITMOS Y DATOS

La noción de *algoritmos* nos conduce a la idea de Inteligencia Artificial (en adelante IA), por lo que importa entender el vínculo entre la IA y los algoritmos,

Mercader Uguina expresa que los datos son el nutriente fundamental de los algoritmos y la inteligencia artificial es el proceso que permite analizar los macrodatos del algoritmo. “La inteligencia artificial (IA) y el aprendizaje automático - agrega el autor español - permiten, a través de esta fórmula, reelaborar gran cantidad de datos simples, de forma que combinados entre sí a través de técnicas estadísticas y lógicas, hacen posible determinar, analizar y predecir ciertos aspectos de la personalidad o el comportamiento, los intereses y los hábitos de una persona”².

La IA no es una inteligencia comparable a la inteligencia humana. Hablamos de IA para indicar “la simulación de procesos de inteligencia humana por parte de máquinas, incluyendo el aprendizaje, el razonamiento y la autocorrección”³.

Aclara esta idea Corvalán, para quien “la inteligencia artificial no es una sola cosa, sino más bien una ‘constelación’ de procesos y tecnologías que permiten que las computadoras complementen o reemplacen tareas específicas que de otro modo serían ejecutadas por seres humanos, como tomar decisiones y resolver problemas”⁴. La IA por lo tanto no opera como la inteligencia humana, sino que “actúa optimizando la ejecución de tareas computarizadas, asignadas por seres humanos, mediante repetición e intentos iterativos”⁵.

Lo que denominamos IA - plantea el economista Rodríguez - es el proceso por el cual introducimos a una computadora los elementos claves del problema a resolver en forma de “datos” que ingresan en su memoria. Estos datos son procesados utilizando algoritmos,

² Mercader Uguina J. R., “Algoritmos y derecho del trabajo” en rev. *Actualidad Jurídica Uría Menéndez*, Madrid 2019, p. 64

³ Sobre la IA, ver: <https://www.gub.uy/agencia-gobierno-electronico-sociedad-informacion-conocimiento/tematica/inteligencia-artificial> , recuperado el 28.03.2023.

⁴ Corvalán J.G, “El impacto de la Inteligencia Artificial en el trabajo”, en *Revista de Direito Econômico e Socioambiental*, V. 10, N° 1, 2019, Curitiba, p. 37.

⁵ *idem*

también ingresados en la computadora y el resultado es una serie de soluciones posibles del problema planteado, donde cada una de ellas tiene un porcentaje de probabilidad de éxito. Esto es lo que se denomina IA y conceptualmente el resultado es similar al dictamen de un humano sobre la mejor acción para solucionar un problema⁶. “En última instancia - afirma - es el histórico dilema de Alan Turing de la década del cincuenta sobre la inteligencia artificial: ¿Puede una computadora comunicarse lo suficiente bien como para persuadir a un humano que es un ser humano?”⁷

EL IMPACTO DE LA IA EN EL TRABAJO

En la línea que venimos señalando, la IA al seleccionar y ordenar datos a través de una lógica algorítmica, se vuelve un instrumento de toma de decisiones que no escapa al sistema de relaciones laborales, abarcando las más diversas áreas desde la selección del personal hasta la forma y modo de ejercicio del poder de dirección, incluyendo la dimensión disciplinaria⁸.

Importa señalar - como lo hace Cañigueral - que hoy los algoritmos y la IA comienzan a intervenir en diversas funciones laborales, independientemente de cuanto más o menos tecnológica sea la empresa⁹. Su impacto va a ser idéntico en la empresa de alta tecnología como en la fábrica tradicional, en que diversas funciones de la organización del trabajo van siendo ocupadas por los nuevos espacios de la inteligencia artificial.

La IA y los algoritmos comienzan por lo tanto a operar en diversos sectores de las relaciones laborales, que obraban habitualmente

⁶ Rodríguez J.M., “Cambio tecnológico y adaptación de la oferta educativa a la nueva demanda de habilidades en el Uruguay”, CEPAL 2020, p. 15

⁷ idem

⁸ Mercader Uguina J. R., ob. cit., p. 63

⁹ Albert Cañigueral, es un referente español sobre el trabajo de plataformas, autor del libro “El trabajo ya no es lo que era”: citado por Gomez M. V., “Los algoritmos llegan a la negociación colectiva”, diario El País, Madrid, 25.04.2021.

bajo la gestión del departamento de personal y/o recursos humanos. Entre ellos señalamos: la selección y contratación de trabajadores, la evaluación de desempeño y la medición de la productividad, la definición de horarios, turnos y tareas, hasta llegar a las medidas disciplinarias y despidos.

Como hemos adelantado, la cuestión “laboral” en torno a los algoritmos se planteará con relación a los datos que alimentan ese proceso de decisión, a la legitimidad de los mismos y al derecho de los trabajadores a conocer los criterios que determinaron la introducción de esos componentes o - como los hemos llamado - “ingredientes” de la receta algorítmica. También está vinculada a la necesidad u oportunidad del uso de programas de gestión por parte de la empresa y a la cuestión de si los mismos pueden lesionar los derechos humanos de la persona, en primer lugar el derecho a la protección de su intimidad y privacidad.

LAS VENTAJAS QUE PROPORCIONA LA LÓGICA ALGORÍTMICA AL QUEHACER HUMANO

En la evaluación de los aspectos positivos vinculados al impacto de la inteligencia artificial en el trabajo humano, una primera ventaja deriva del hecho que las operaciones algorítmicas actúan con mayor rapidez y objetividad que la lógica humana. El ser humano al seleccionar y decidir, inevitablemente actúa con una subjetividad propia de su condición. En cambio, las decisiones algorítmicas “no se basan en corazonadas, imprecisiones o prejuicios”¹⁰.

La IA, al ordenar y descubrir vínculos entre datos, selecciona generalmente las mejores opciones: como consumidores hemos

¹⁰ Pasquale F.A., “The Black Box Society: The Secret Algorithms that control money and information, Cambridge”, Harvard University Press, 2015, p. 35, citado por Mercader Guina J. R., ob. cit., p. 64

aprendido a elegir en función de evaluaciones algorítmicas, ya se para reservar un hotel en el exterior o para adquirir un producto en *Mercado Libre* o en *Amazon*: el emoticón “pulgar para arriba”, es decir “me gusta”, condiciona muchas de nuestras decisiones a la hora de tener que optar.

Panizza señala que las nuevas tecnologías informáticas, además de promover el uso de *tablets* y *smartphones*, permiten construir equipos de control remoto, sensores de seguridad, simuladores de tareas de todo tipo, relojes inteligentes, mientras drones de diverso tipo realizan inventarios, trasladan desfibradores al lugar de trabajo, aplican plaguicidas en zonas rurales. Además, la acumulación de datos (*big data*) contribuye a la resolución de los problemas generados en el trabajo y la producción¹¹.

Gauthier expresa que los ordenadores son buenos para ciertos tipos de tareas, tales como la identificación de errores, procesamiento de datos y el cálculo de cifras financieras. Sin embargo, son menos capaces de realizar otras funciones como reconocer la ironía, determinar si algo es NSFW (*no seguro para el trabajo* en inglés) o discernir entre los resultados de búsquedas ambiguas¹².

Es importante también destacar que, como ha señalado la Agencia Europea para la Seguridad y Salud en el Trabajo, estos nuevos sistemas pueden facilitar la comprensión de las alternativas relativas a la seguridad y salud en el trabajo, la toma de mejores decisiones y la predicción de los problemas en materia preventiva antes de que ocurran, así como pueden favorecer intervenciones más oportunas y eficaces. Incluso pueden facilitar que las

¹¹ Panizza C., “Impacto de las nuevas tecnologías en la salud”, exposición en el marco del taller *Impacto de las nuevas tecnologías en los derechos laborales fundamentales*, organizado por la Asociación Uruguaya de Derecho del Trabajo y la Seguridad Social, Montevideo 24.06.2021.

¹² Gauthier G., “Economía compartida, *crowdworking* y Derecho del trabajo”, en AA.VV., “Disrupción, Economía Compartida y Derecho”, FCU, Montevideo, junio de 2016, p. 113.

inspecciones laborales investiguen de forma más eficaz los eventuales incumplimientos¹³.

LOS “PROBLEMAS” DE LA IA Y EL ALGORITMO

El impacto de la dimensión digital, representado por la IA y los algoritmos, produce sin embargo problemas diversos, no siempre fáciles de solucionar. Por otra parte, la rapidez de la evolución tecnológica determina la aparición de complejidades nuevas, que no se resuelven con la misma velocidad. Algunos de estos problemas aterrizan con fuerza en las relaciones laborales. Señalaremos los siguientes en una enumeración, por supuesto, no taxativa:

PROBLEMA: LA “ACCESIBILIDAD” AL ALGORITMO

Como expresáramos, la IA toma sus decisiones a partir de la información acumulada en los algoritmos, la cual a su vez responde a un previo diseño humano de conseguir un logro o alcanzar determinado resultado. Como señalábamos en el ejemplo de la torta, la calidad y determinación del producto o decisión algorítmica dependerá de los “ingredientes” - es decir, los datos - con los cuales alimentamos el proceso.

Esto datos no son fáciles de identificar para el hombre común y uno de los desafíos de los próximos tiempos será precisamente el de dar claridad y comprensión a los insumos que alimentan el algoritmo, para que el trabajador y su organización puedan acceder a ellos de una forma comprensible.

¹³ Mercader Uguina J.R., ob. cit., p. 68, quien cita *Estudio prospectivo sobre riesgos nuevos y emergentes para la seguridad y salud en el trabajo asociados a la digitalización en 2025* (Oficina de Publicaciones de la Unión Europea, Agencia Europea para la Seguridad y la Salud en el Trabajo, 2018)

PROBLEMA: LA DISCRIMINACIÓN CIENTÍFICAMENTE CONCEBIDA

La noción de trabajo decente implica no ser objeto de ningún tipo de discriminación. El problema que identificamos en la lógica artificial se vincula al hecho que es posible producir de modo oculto efectos discriminatorios en el momento de armar la programación del algoritmo. La acción arbitraria se lleva a cabo, introduciendo o eliminando datos en la “receta” del algoritmo (relativos al género, la raza, la edad, el origen nacional, etc.), que alteran la igualdad entre los trabajadores. Los datos que alimentan el algoritmo no son necesariamente visibles para la persona afectada y la discriminación puede producirse al comienzo, es decir durante el proceso de selección de los trabajadores; durante la ejecución del contrato de trabajo evaluando con criterios distintos a personas en igualdad de condiciones; o a la finalización del vínculo laboral, favoreciendo el egreso de unos y no de otros en función de determinadas características personales y no de su capacidad laboral.

PROBLEMA: LA INVASIÓN DE LA VIDA PRIVADA Y FAMILIAR DEL TRABAJADOR

El uso de los datos personales para la construcción de los algoritmos permite ingresar a aspectos de la vida privada, que exceden el vínculo funcional y pueden poner en riesgo el derecho a la privacidad no solo del trabajador, sino también de su núcleo familiar. Por lo tanto debe existir una adecuada protección en materia de comunicación de datos de carácter personal de las personas que conviven con el trabajador, respeto a la diversidad cultural, religiosa y lingüística de ese ámbito tan especial. Debe asegurarse en tales casos el derecho a la intimidad no solo personal, sino también familiar, evitando cualquier intromisión

ilegítima que reproduzca en datos la vida íntima de las personas en lugares o momentos de su vida privada¹⁴.

PROBLEMA: LA PRESIÓN SOBRE LA PRODUCTIVIDAD DEL TRABAJO

La organización algorítmica del trabajo puede llevar a presionar de forma “deshumana” a los trabajadores, para que aumenten su productividad. Ya no será un capataz o un supervisor, quien controle el volumen y la rapidez del trabajo de cada trabajador, sino un sistema tecnológico, que no tiene momentos de distracción o de pausa. Ello llevará inevitablemente a situaciones de stress y daños a la salud, cuyo estudio comienza a ocupar y preocupar a especialistas en tema de salud psicosocial.

PROBLEMA: EL CONTROL ALGORÍTMICO

La recolección de datos del trabajador implicará una nueva modalidad de vigilancia digital, sumándose a ello el empleo de sensores que permiten un control de emociones y movimientos. A modo de ejemplo, la ropa “wearable” (“ropa ponible” en su significado literal, pero que acostumbramos a traducir como “ropa inteligente”) constituye un nuevo modelo de recopilación de datos del trabajador. Iglesias señala que en la noción de ropa inteligente y otros accesorios (wearables) ingresan elementos como calzados, lentes, cascos, chaquetas tecnológicas, relojes y pulseras inteligentes, cada vez más difundidos en los más diversos sectores de la actividad industrial y de servicio¹⁵. La autora alerta sobre el

¹⁴ Entre las previsiones jurídicas que tutelan el derecho a la privacidad del núcleo familiar, señalamos el art. 8 de la Convención Europea de Derechos Humanos, que establece el derecho de toda persona al respeto de su vida privada y familiar, de su domicilio y de su correspondencia.

¹⁵ Iglesias L., “Impacto de las nuevas tecnologías en los derechos humanos fundamentales”, exposición en el marco del taller *Impacto de las nuevas tecnologías en los derechos laborales fundamentales*, organizado por la

riesgo que las mismas significan en relación a la afectación de la privacidad del trabajador y entiende que su utilización debe estar guiada por específicos criterios como la estricta pertinencia para la función, una información clara y veraz sobre los fines perseguidos, que no impliquen un control sobre la vida personal y no se apliquen en forma clandestina sobre el trabajador¹⁶.

LAS NORMAS Y PRINCIPIOS TUTELARES ANTE LOS ALGORITMOS

Ante el avance vertiginoso de las nuevas tecnologías y los progresos de la automatización y la inteligencia artificial, muchas veces el Derecho del trabajo reacciona con lentitud, por lo cual nuestro desafío es ir construyendo un nuevo derecho a partir de los principios y las soluciones que brinda la experiencia comparada.

Debemos buscar un primer conjunto de tutelas, en aquellas reconocidas por los principios del Derecho del trabajo, las normas constitucionales, los grandes pactos internacionales ratificados por nuestro país y las Declaraciones universales. Serán aplicables a las diversas situaciones que pueden afectar al trabajador en el uso ilegítimo o desmedido de los algoritmos, principios y normas relativos a

- Derechos inespecíficos, es decir derechos atribuidos al trabajador en su calidad de ciudadano, como el derecho a la intimidad, privacidad, honor, imagen, etc.
- Principios y derechos vinculados a la integridad psicofísica, como protección de la salud y seguridad ante la siniestralidad en el trabajo

Asociación Uruguaya de Derecho del Trabajo y la Seguridad Social, Montevideo
24.09.2022

¹⁶ *idem*

- Principio de la igualdad y no discriminación (género, raza, edad, orientación sexual, etc);
- Normas contra la violencia en el trabajo, en especial aquellas relativas a las diversas formas de acoso.

En una segunda línea importa conocer la evolución del derecho comparado, que adquiere especial importancia en consideración a la rápida evolución de la realidad. Las reglas consignadas en normas extranjeras o en documentos internacionales no suscritos por nuestro país (como los que conforman el derecho comunitario de la Unión Europea) pueden perfectamente ingresar a nuestro derecho positivo, como doctrina más recibida, sin perjuicio que los textos aprobados en otros países o regiones puedan constituir un referente para el legislador en oportunidad de discutir y aprobar reglas para la realidad nacional. Importa por lo tanto destacar algunos cuerpos normativos que consideramos de estudio indispensables para afrontar los nuevos temas y problemas de la IA y el trabajo.

Señalamos como expresión fundada del derecho comparado internacional el “Reglamento General de Protección de Datos del Parlamento Europeo” (RGPD¹⁷), que entró en vigor el 24/05/2016, con aplicación a partir del 25/05/2018. El documento estableció ese “período ventana de dos años”, para que las empresas, las organizaciones, los organismos y las instituciones de los países de la Unión, pudieran adaptarse a las nuevas reglas, para efectivizar su cumplimiento.

El Reglamento establece los siguientes principios generales en su art. 5:

- Los datos personales deben ser tratados de forma lícita, leal y transparente.

¹⁷ en inglés GDPR (*General Data Protection Regulation*)

- Los datos personales deben ser recogidos con fines determinados, explícitos y legítimos.
- Los datos personales deben ser adecuados, pertinentes y limitados a lo necesario en relación con el tratamiento.
- Los datos personales deben ser exactos y estar siempre actualizados.
- Los datos personales deben mantenerse de forma tal, que se permita la identificación de los interesados durante no más tiempo del necesario para los fines del tratamiento.
- Los datos personales deben ser tratados de tal manera que se garantice su seguridad.

El Reglamento constituye una normativa a nivel de la Unión Europea, por lo que cualquier empresa de la unión, o aquellas empresas que tengan negocios en la Unión Europea y manejen información personal de cualquier tipo sobre sus trabajadores, deben acogerse a ella. Las sanciones son graves y se prevén multas por el no cumplimiento del RGPD, que pueden llegar a los 20 millones de euros.

Una tercera vía de estudio obliga a conocer la más actualizada doctrina extranjera de países, cuyos sistemas de relaciones laborales posean características análogas al nuestro. A tal propósito Rojo Torrecillas releva una serie de libros y publicaciones de la actual doctrina española e internacional, que es bueno consignar en este trabajo, para quienes quieran ampliar su lectura¹⁸.

¹⁸ Rojo Torrecillas en su blog del 28/08/2021, http://www.eduardorojotorrecilla.es/2021/08/tecnologia-y-relaciones-laborales_28.html, recuperado el 07/03/2023 señala diversos libros, artículos y ponencias - en especial de la doctrina española - sobre algoritmos e inteligencia artificial:: Pilar Rivas, "La aplicación de la inteligencia artificial al trabajo y su impacto discriminatorio", 2021; Henar Álvarez Cuesta, Antonio José Valverde,

EL DERECHO A LA TRANSPARENCIA ALGORÍTMICA

En la Unión Europea va imponiéndose el derecho a la transparencia algorítmica, que se expresa en la posibilidad que la persona - en nuestro caso, el trabajador - cuyos datos sean utilizados para construir algoritmos y por lo tanto ser alcanzado por decisiones de la Inteligencia Artificial, pueda tomar conocimiento de como se elaboraron los mismos: en el sencillo ejemplo que hemos empleado en este artículo, es el derecho del trabajador a conocer la “receta de cocina”, que de una u otra forma tendrá consecuencias sobre su persona.

“Implantación de sistemas de inteligencia artificial y trabajo”, 2020; Francisco Javier Fernández Orrico, “Criterios sobre uso de dispositivos tecnológicos en el ámbito laboral. Hacia el equilibrio entre el control empresarial y la privacidad del trabajador”; Jesús Cruz Villalón, “Transformación digital y relaciones laborales”; María Luz Rodríguez Fernández, “Algorithms and Robots: The challenges of working in the digital age”; José Luis Goñi Sein, “El impacto de las nuevas tecnologías disruptivas sobre los derechos de privacidad (intimidad y extimidad) del trabajador”; Cristóbal Molina Navarrete, “Duelo al sol (digital). ¿Un algoritmo controla mi trabajo? Sí; a tu empresa también”; Antonio Álvarez Montero, “Más allá de los sectores “punteros”. La negociación de contenidos digitales se extiende a empresas tradicionales”; Sonia Fernández Sánchez, “Frank, el algoritmo consciente de Deliveroo. Comentario a la sentencia del Tribunal de Bolonia 2949/2020 de 31 de diciembre”; Margarita I. Ramos Quintana, “La violencia de género en las relaciones laborales: manifestaciones diversas anteladas nuevas formas de prestar trabajo”; Ana Belén Muñiz Ruiz, “¿Damos suficiente cabida a la emoción en el modelo laboral digital? ; Daniel Pérez del Prado, “El debate europeo de los algoritmos”; Jesús R. Mercader en “Discriminación algorítmica en el trabajo y derecho fundamental a la transparencia: ¿Debemos (podemos) regular los algoritmos?”; José María Goerlich, “Algoritmos en el procedimiento sancionador en el orden social y en la gestión recaudatoria de la Seguridad Social: las decisiones automatizadas en el Real Decreto-ley 2/2021”; Ana M^a de la Puebla, “El RDL 2/2021, de 26 de enero: nueva entrega de la prórroga de los ERTE y el aterrizaje del Big Data en la ITSS”; Andrea Lassandari, “La tutela collettiva del lavoro nelle piattaforme digitali: gli inizi di un percorso difficile”. Agrego de Serrani L. y Tiraboschi M. (Coordinadores), “El sistema de protección social y prevención de riesgos laborales frente a la IV Revolución Industrial”, 2021.

Este derecho es reconocido *in primis* por el propio RGPD, que regula en su art. 12 la “*Transparencia de la información, comunicación y modalidades de ejercicio de los derechos del interesado*” y expresa que el responsable del tratamiento de los datos deberá tomar las medidas oportunas para facilitar al interesado toda información relativa al uso y tratamiento de sus datos personales. Esta información deberá ser comunicada “en forma concisa, transparente, inteligible y de fácil acceso, con un lenguaje claro y sencillo, en particular cualquier información dirigida específicamente a un niño. La información será facilitada por escrito o por otros medios, inclusive, si procede, por medios electrónicos”. Agrega la norma que cuando la solicite el interesado, la información podrá facilitarse verbalmente siempre que la persona demuestre por otros medios su identidad.

A su vez, el art. 15 reconoce el derecho de acceso a la información, estableciendo que el interesado tendrá derecho a obtener del responsable del tratamiento confirmación de si se están tratando o no datos personales y, en tal caso, tener derecho de acceso a la información relativa a: los fines del tratamiento, las categorías de datos personales de que se trate, los destinatarios, el plazo previsto de conservación de los datos personales la rectificación o supresión de datos personales o la limitación del tratamiento de datos personales, o a oponerse a dicho tratamiento y el derecho a presentar una reclamación ante una autoridad de control.

También destacamos el art. 21 del RGPD que establece el derecho de la persona a oponerse a que sus datos personales sean objeto de un tratamiento para la elaboración de perfiles. Por su parte, el art. 47, lit. j) establece que los responsables de la utilización de los datos deberán llevar adelante auditorías de protección de dichos datos y métodos para garantizar acciones correctivas con la finalidad de proteger los derechos del interesado.

En materia de derecho a la transparencia debe también destacarse el Real Decreto ley español N° 9/2021, que introduce en el Estatuto de los Trabajadores el derecho de los Comités de Empresa a “ser informados por la empresa de los parámetros, reglas e instrucciones en los que se basan los algoritmos o sistemas de inteligencia artificial que afectan a la toma de decisiones que pueden incidir en las condiciones de trabajo, el acceso y mantenimiento del empleo, incluida la elaboración de perfiles”.

Importa destacar que el convenio colectivo del sector financiero citado en el numeral anterior, en el art. 35, num. 5 - entre los derechos del trabajador ante la IA - reconoce el derecho de información sobre la analítica de datos y la lógica de funcionamiento de los algoritmos. La norma expresa que la empresa deberá informar a la representación de trabajadores (*RLT*, sigla jurídica de *Representación Legal de Trabajadores*) sobre el uso de la analítica de datos o los sistemas de inteligencia artificial cuando los procesos de toma de decisiones en materia de recursos humanos y relaciones laborales se basen exclusivamente en modelos digitales sin intervención humana. Dicha información, como mínimo, abarcará los datos que nutren los algoritmos, la lógica de funcionamiento y la evaluación de los resultados¹⁹.

CONCLUSIONES: LOS DESAFÍOS FUTUROS DE LAS RELACIONES LABORALES

Las transformaciones tecnológicas del Siglo XXI han tenido un impacto profundo no solo en la forma de trabajar y de organizar la producción, sino en la propia estructura de las empresas que han pasado en pocos años de una dimensión real y “física” a un formato virtual, en el que la gestión digital del trabajo desempeña un rol protagónico.

¹⁹ BOE de 15/10/2021 cit.

La rapidez de los cambios obliga a estudiar un fenómeno de muy difícil aprehensión desde la perspectiva de una realidad intangible y que por la rapidez en los cambios, no logra ser regulada suficientemente y a tiempo por el legislador.

El modelo de empresa virtual ha experimentado un crecimiento en tiempos de la pandemia del COVID 19, porque ésta ha sido el laboratorio perfecto a nivel global para experimentar y desarrollar todo tipo de tecnología digital. Al mismo tiempo, la tecnología no solo promueve la expansión de actividades como el e-commerce, las diversas formas de teletrabajo, la automatización de muchos procesos productivos, sino que ingresa en el núcleo duro de la empresa, para organizar su gestión a través de procesos algorítmicos, que englobamos bajo el nombre de Inteligencia Artificial y que confían a la propia IA decisiones sobre la contratación, supervisión y extinción del contrato de trabajo.

La transformación digital de la empresa obliga a nuevos vínculos entre humanos y e IA, para los cuales se emplea la nueva expresión de *cobotización* (sic), que refiere a la convivencia de humanos con máquinas inteligentes y robots. El riesgo es que legisladores y organizaciones de trabajadores no presten rápidamente atención a esta realidad, en la que es necesario establecer reglas claras de colaboración entre personas y máquinas, evitando que los algoritmos ocupen cada vez mayores espacios en procesos de decisión, tradicionalmente ejecutados por humanos.

En especial preocupa comprobar - desde nuestra experiencia - que las organizaciones sindicales no han asumido un rol activo ante las complejidades de este nuevo modelo virtual de gestión del trabajo. Entendemos que los sindicatos consideran que las nuevas tecnologías son peligrosas para los trabajadores (lo cual es una percepción correcta), pero no reaccionan en forma adecuada, sino que optan por ignorar estos cambios, pensando que la ignorancia de las nuevas realidades reduce los problemas. Por tal motivo es pobre la negociación colectiva sobre las nuevas tecnologías e

inexistente la resistencia ante el embate de los algoritmos en la gestión del trabajo, mientras los convenios colectivos en la materia siguen siendo una asignatura pendiente en América Latina.

Uno de los temas por lo tanto que más deberá ocupar la atención de las organizaciones y de los especialistas en relaciones laborales es la necesidad que los trabajadores puedan establecer una barrera de tutelas ante la una lógica digital, que a pretexto de objetivizar las decisiones de la empresa, oculte datos que quiebren principios como el de igualdad, privacidad y respeto de los derechos humanos. En esa acción, que involucra la responsabilidad de académicos, legisladores y actores sociales, deberán promoverse reglas que permitan tener acceso a la información que alimenta el algoritmo, para entender su lógica subyacente y denunciar eventuales ilegitimidades en la composición de los datos aportados.

También importa señalar que los jueces laborales tendrán un rol central en la metabolización de las normas y los principios aplicables a la intervención de la IA en el contrato de trabajo, así como en la definición de los criterios de prueba, para analizar realidades nuevas y en las que probablemente tendrá especial destaque la prueba pericial.

En definitiva, el gran desafío y el objeto del debate global será el de saber ajustar el derecho del trabajo al fenómeno de la gestión digital del trabajo, que deja el trabajador cada vez más expuesto - en su vida laboral y personal - a los controles y a las presiones de decisiones automáticas, realizadas por procesos algorítmicos, desde el ámbito de la empresa.

NOVAS PERSPECTIVAS PARA O CONTROLE EXTERNO: ONLINE DISPUTE RESOLUTION, IA E TRIBUNAIS DE CONTAS

Ricardo Schneider Rodrigues.

Pós-doutorando em Controle Externo e Novas Tecnologias pelo Grupo de Pesquisas SmartCitiesBr-EACH da Universidade de São Paulo (USP). Doutor em Direito pela Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS). Mestre em Direito Público pela Universidade Federal de Alagoas (Ufal). Procurador do Ministério Público de Contas de Alagoas.

INTRODUÇÃO

As revoluções industriais vêm produzindo significativas mudanças ao longo da história. Desde a Indústria 1.0, com o advento da máquina a vapor, passando pela Indústria 2.0, com a máquina a diesel e a eletricidade, pela Indústria 3.0, com sistemas computadorizados, até os dias atuais, com a Indústria 4.0, em que emergem novas tecnologias, como a inteligência artificial (IA), a impressora 3D, a internet das coisas (*Internet of Things - IoT*), a nanotecnologia, os veículos automatizados e a computação quântica, a sociedade tem passado por grandes transformações (RISTIANDY, 2020, p. 87-88).

As mudanças nas relações estabelecidas entre os indivíduos, em diversas esferas, são incontáveis. As novas tecnologias de comunicação e informação propiciam a transição de várias temáticas para as vias virtuais, estimulando a disseminação de significativas inovações na estrutura social (CASTELLS, 2011).

As novas relações de trabalho, muitas delas agora intermediadas por aplicativos, alteram os modos até então usuais de prestação de

serviços e suscitam diversos questionamentos, ainda mais quando se tenta analisá-las a partir da perspectiva de modelos econômicos e de institutos jurídicos tradicionais (MARQUES; FREITAS, 2017).

A possibilidade de processar e analisar uma grande quantidade de dados (*big data*), pelo uso da IA, permite obter informações pessoais que dificilmente seriam alcançáveis caso fossem tratadas individualmente e por métodos tradicionais. A preocupação com a proteção de dados pessoais, cada vez mais usados como moeda de troca para viabilizar o acesso a quase todo tipo de serviços, é crescente (DONEDA, 2011).

A Administração Pública não está imune a esse contexto. Embora os reflexos dos avanços da tecnologia e da inovação tendam a repercutir de forma mais lenta no âmbito do poder público, não há na sociedade atual espaço impermeável a tais mudanças.

A velocidade de tais avanços impõe ao poder público, em especial aos órgãos de controle, atentar para os possíveis usos dessas novas tecnologias como forma de aprimorar o desenvolvimento de suas atividades. O foco do presente trabalho recai sobre o uso das novas tecnologias para fins de controle da Administração Pública, em especial aquele desempenhado pelos Tribunais de Contas.

Pretende-se avaliar se o uso a inteligência artificial (IA) poderia maximizar a atuação dessas Cortes, especialmente no atual contexto em que tecnologias como o ChatGPT, desenvolvido pela OpenAI, já avançaram sobremaneira no desenvolvimento da capacidade de utilizar grandes bases de dados (*big data*), realizar análises e produzir textos emulando a linguagem natural de forma muito aproximada à produzida pelo homem.

Dada a vastidão de novidades e as mudanças que surgem a cada dia, o recorte epistemológico proposto busca delimitar a análise empírica e propositiva desta pesquisa ao exame da pertinência do desenvolvimento de plataformas semelhantes às utilizadas no

âmbito privado para a Resolução *Online* de Controvérsias ou Litígios (*Online Dispute Resolution - ODR*), porém adaptadas às especificidades dos Tribunais de Contas.

A TRANSFORMAÇÃO DIGITAL NO SETOR PÚBLICO: A LEI DO GOVERNO DIGITAL E OS TRIBUNAIS DE CONTAS

Em todo o mundo, observa-se um movimento dos governos no sentido de alcançar a transformação digital (*Digital Transformation - DT*), vista como um imperativo estratégico para aprimorar a prestação do serviço, a experiência dos “clientes”, otimizar operações e criar novos modelos de negócio. Exemplos de estratégias de *DT* podem ser observados na União Europeia, com o “*EU e-Government Action Plan 2016-2020*”, nos Estados Unidos, com a “*Digital Government Strategy*”, e na China, com o programa “*Internet Plus Government Services*” (GONG; YANG; SHI, 2020).

O Governo da Província de Zhejiang (*Zhejiang 1Provincial Government - ZPG*), na China, é um exemplo disso. Abrangendo três diferentes níveis governamentais - uma província, 11 prefeituras e 89 condados, além de 42 departamentos -, e uma população de 57 milhões de habitantes (no final de 2018), o ZPG começou a entregar serviços *online* baseados em nuvem em 2014. Depois, usou as tecnologias de *big data* e é pioneiro na *transformação digital* do governo digital naquele país. Lá foram identificadas três ondas, resultando, na última fase, na entrega *online* de 95,2% dos serviços administrativos, na transformação de 75% dos serviços interdepartamentais em serviços de balcão único, na redução em 29% do número de documentos necessários para submissões e em 35% do tempo de processamento, em comparação com a sistemática anterior (GONG; YANG; SHI, 2020).

Na Indonésia, o *e-government* vem se expandindo cada vez mais, de modo que o Presidente chegou a declarar que os terceiro e quarto escalões seriam substituídos pela IA. O uso do *e-budgeting* no país, sistema de planejamento orçamentário realizado de forma eletrônica, já teria propiciado a redução da malversação do orçamento público em províncias como Jakarta e Surabaya. Outra iniciativa relacionada a essas mudanças é a criação pelo governo indonésio de um cartão de identificação integrado, conhecido como *Eletronic KTP (E-KTP)*, por meio do qual passou a prestar serviços com efetividade e eficiência (RISTIANDY, 2020).

Para Ristiandy, estaria em curso na Indonésia uma disrupção burocrática, com mudanças significativas em diversos aspectos, como nos procedimentos até então adotados, na forma de prestar serviços públicos, na forma de recrutamento de pessoal e nas vagas a serem ofertadas futuramente. Para o autor, há a expectativa de que o *e-government* modifique a forma como a população enxerga o funcionalismo, ao reduzir a corrupção, o conluio e o nepotismo no país (2020).

Na perspectiva ambiental, Putrijanti defende que o governo utilize a revolução industrial 4.0 para administrar a própria revolução, de forma a melhorar a qualidade do meio ambiente e da sociedade. Propõe o uso da tecnologia da informação para controlar o desperdício e regular a participação da sociedade na tomada de decisão (2019).

É preciso compreender essa nova realidade, para adotar uma postura propositiva capaz de contribuir com mudanças reais e positivas. Nesse sentido, na perspectiva brasileira, o ponto de partida deve recair sobre a Lei nº 14.129/2021, que dispõe sobre princípios, regras e instrumentos para o Governo Digital (*e-government*) e para o aumento da eficiência pública. A referida norma é aplicável expressamente ao Tribunal de Contas da União (TCU) e, mediante regulamentação específica, aos demais

Tribunais de Contas (TCs) do país (art. 2º, inc. I e III, § 2º, Lei nº 14.219/2021).

Entre seus comandos, o referido normativo prevê, como princípios, diversas diretrizes voltadas ao aperfeiçoamento do controle da Administração Pública. Segundo dispõe a nova lei, deve existir uma nova relação entre o poder público e a sociedade, pautada por desburocratização, modernização, fortalecimento e simplificação, mediante serviços digitais, cujo acesso deve ocorrer inclusive por dispositivos móveis (art. 3º, inc. I).

À luz do novo regramento, o acesso a serviços públicos e informações deve se dar, se possível, em plataforma única, permitindo aos cidadãos, pessoas jurídicas e até outros entes públicos demandar e acessar serviços públicos digitalmente, dispensando a exigência de solicitação presencial (art. 3º, incisos II e III). Os sistemas devem ser pautados pela interoperabilidade e pela promoção de dados abertos (art. 3º, inc. XIV). Os procedimentos de solicitação, oferta e acompanhamento dos serviços públicos devem ser simplificados, universalizados e, se possível, viabilizados por meio do autosserviço (art. 3º, inc. X).

Além disso, a lei prevê a implantação do governo como plataforma e promove o uso de dados, preferencialmente anonimizados, por pessoas físicas e jurídicas, de diversos setores da sociedade, resguardadas as diretrizes da Lei nº 13.709/2018 (LGPD), para, entre vários fins, em especial, aperfeiçoar a formulação de políticas públicas e o controle social (art. 3º, inc. XXIII).

A participação social no controle e na fiscalização da Administração Pública, neste novo contexto, de um Governo Digital, passa a ser incentivada, e ao gestor público atribui-se o dever de prestar contas sobre a gestão dos recursos públicos diretamente à população (art. 3º, incisos IV e VI). É prevista a possibilidade de atuação integrada e de compartilhamento de dados pessoais, em ambiente seguro, quando indispensável, e até

a transferência de sigilo, entre órgãos e entidades envolvidos na prestação e no controle de serviços públicos, observadas as disposições da LGPD, da Lei nº 5.172/1966 (CTN) e da Lei Complementar nº 105/2001 (art. 3º, inc. IX).

Passa a ser preferencial, no uso da internet e de suas aplicações, a adoção de tecnologias, de padrões e de formatos abertos e livres, nos termos do Marco Civil da Internet, como previsto na Lei nº 12.965/2014 (art. 3º, XXV). O desenvolvimento tecnológico e a inovação no setor público passam a ser princípio e diretriz da Administração Pública (art. 3º, inc. XXVI).

Como visto, os ditames da Lei do Governo Digital alcançam o Tribunal de Contas da União, órgão de controle externo que é modelo para as demais Cortes de Contas do país, nos termos do art. 75, *caput*, da Constituição e de diversos julgados do Supremo Tribunal Federal (STF).¹

Nesse sentido, a Associação dos Membros dos Tribunais de Contas do Brasil (Atricon) aprovou a Nota Recomendatória Conjunta Atricon/IRB/Abracon/CNPTC/Audicon nº 2/2022, que propõe aos Tribunais de Contas brasileiros a adoção dos princípios, regras e instrumentos da Lei do Governo Digital, para estimular a adesão por parte dos seus jurisdicionados (2022).

Constatada a existência de um marco jurídico hábil a induzir a transformação digital no setor público, alcançando inclusive os TCs, cumpre avaliar o cabimento do uso da ODR e da IA como instrumentos capazes de aprimorar o controle externo da Administração Pública.

¹ O STF tem jurisprudência pacífica quanto à obrigatoriedade de os Tribunais de Contas Estaduais seguirem o modelo federal (TCU) de organização, composição e fiscalização, inclusive em relação ao Ministério Público de Contas, nos termos do art. 75 da Constituição (princípio da simetria). Nesse sentido, cf. ADI 4.416 MC, ADI 3.307, ADI 916 e ADI 3.715 MC.

O USO DA ONLINE DISPUTE RESOLUTION PELOS TRIBUNAIS DE CONTAS: UM AMBIENTE PROMISSOR PARA O USO EFETIVO DA INTELIGÊNCIA ARTIFICIAL

Não se desconhece o uso de iniciativas importantes no âmbito dos Tribunais de Contas, voltadas ao uso da inteligência artificial (OLIVEIRA, 2022; GONZALEZ, 2023). Sem embargo, pretende-se avaliar neste trabalho a possível aplicação da IA em um instrumento específico, amplamente utilizado no âmbito privado e com potencial de utilização também pelo setor público: a *Online Dispute Resolution* (AMORIM; RODRIGUES, 2019, p. 174).

A ODR consiste no uso das tecnologias da informação e da comunicação (*Information and Communication Technologies - ICT*) para apoiar a resolução de conflitos em ambiente virtual. Surgiu como uma extensão *online* da Resolução Alternativa de Litígios (*Alternative Dispute Resolution - ADR*), que adota como métodos a negociação, a conciliação, a mediação e a arbitragem, alcançando sucesso na composição de disputas *offline*. A *ADR* surgiu como uma opção privada ao congestionamento, aos custos, à demora e ao formalismo do Judiciário (AMORIM; RODRIGUES, 2019).

Arbix destaca que não se trata apenas da substituição de mecanismos *offline* por *online*. Aspecto central, na visão do autor, corresponde ao papel interativo que as novas tecnologias exercem, impactando sobre os ambientes e procedimentos em que atuam, alterando sua dinâmica tanto na autocomposição, pelo uso de métodos consensuais, quanto na heterocomposição (adjudicação). Em outros termos, a *ODR* não se limita a ser uma *ADR online*, quer pelo papel ativo que a tecnologia exerce na solução da controvérsia - não basta que as formas de resolução de conflitos sejam apenas atualizadas pela tecnologia, mas que modifiquem os ambientes e procedimentos, atuando com uma

“quarta parte” –, quer por abranger situações outras além da consensualidade (ARBIX, 2017, p. 57-65).

Esses dois aspectos são relevantes para os propósitos deste trabalho, pois a avaliação do uso da *ODR* no âmbito dos TCs pretende examinar como essa ferramenta poderia ser adaptada a esse contexto específico, para que a tecnologia seja uma ferramenta útil e inovadora tanto em busca de soluções consensuais no âmbito da Administração Pública, como também para propiciar uma solução mais célere e eficaz nos casos em que isso não for possível.

Por tais razões, adota-se neste trabalho o seguinte conceito operacional de *ODR*:

A *ODR* é a resolução de controvérsias em que tecnologias da informação e comunicação não se limitam a substituir canais de comunicação tradicionais, mas agem como vetores para oferecer às partes ambientes e procedimentos ausentes em mecanismos convencionais de dirimir conflitos. (ARBIX, 2017, p. 58).

Há diversas plataformas desenvolvidas para proporcionar a resolução *online* de controvérsias. Desde as primeiras, voltadas ao comércio eletrônico, como o exemplo clássico do Ebay, até as mais recentes, utilizadas no âmbito do Poder Judiciário, como a Modria, em especial a desenvolvida para o *Civil Resolution Tribunal (CRT)*, da Colúmbia Britânica, no Canadá (MALONE; NUNES, 2022, p. 203-235). No Brasil, embora não haja plataformas tão avançadas em operação, como a Modria, há vários exemplos em uso, como a MOL e a Consumidor.gov.br (LOSS; ARBIX, 2022).

O exemplo da *CRT* é importante, pois o caráter “judicialiforme” dos processos de contas permite algumas aproximações. Por tal razão, é importante conhecer, em linhas gerais, sua estrutura de funcionamento. A *CRT* opera em quatro fases. Na primeira, por

meio da ferramenta *Solution Explorer*, auxilia as partes a compreenderem seus problemas e se há, de fato, alguma reivindicação válida e qual o procedimento adequado, disponibilizando documentos a serem utilizados pelos usuários. Na segunda fase, ocorre a negociação automatizada, gerenciada pela própria plataforma, por meio de linguagem pré-estruturada, e o uso de *software* que identifica o problema e oferece soluções. Apenas em caso de não se alcançar acordo nas fases iniciais, que transcorrem sem a necessidade de uma parte “humana”, a controvérsia avança para a terceira fase, em que há um facilitador, que apoia a busca por uma resolução da controvérsia e, se for o caso, para a fase final, em que haverá a decisão do Tribunal, com força de título executivo e passível de recurso (MALONE; NUNES, 2022, p. 215-222; CRT, 2023).

É perceptível, destarte, a possibilidade de se recorrer à inteligência artificial para auxiliar no desenvolvimento de cada etapa da ODR. Nas primeiras fases, sem a presença humana, a IA auxiliará a identificar o problema e a sugerir um desfecho para o caso. Nas demais, com a presença humana, pode apoiar a atuação do facilitador e do julgador.

Susskind, ao tratar dos Tribunais *online*, sugere algo semelhante, em termos de fases a serem implementadas por plataformas judiciais. A primeira, para o autor, teria o papel preventivo de orientação, triagem e avaliação dos casos. A segunda contaria com facilitadores, para conciliação e mediação sem intervenção judicial, que, numa segunda geração, seriam substituídos por ferramentas de negociação *online* e máquinas com sistema de aprendizagem. Por fim, a resolução ocorreria por juízes humanos, em ambiente *online*, que, numa segunda geração, seriam auxiliados pela IA (MALONE; NUNES, 2022, p. 80-86)

Portanto, é relevante o papel conferido à IA nessas plataformas, seja identificando adequadamente a controvérsia, seja sugerindo possíveis acordos aos envolvidos, seja, ainda, apoiando humanos

no desenvolvimento dos seus papéis de conciliadores, mediadores ou julgadores. Com o avanço da tecnologia, essa integração tende a se expandir.

Para Stuart Russel, “uma entidade é inteligente à medida que faz o que provavelmente serve para atingir o que ela quer, levando em conta o que ela percebeu” (2021, p. 23). Desse modo, em termos de inteligência artificial, é crucial assegurar que os objetivos dessas máquinas sejam os mesmos que os nossos, a fim de evitar distorções e outros problemas relevantes, como a questão dos vieses, que não será aprofundada neste trabalho.

Sem embargo, é necessário avançar, ao menos, na compreensão de que a IA pode ser classificada de várias formas e há vários graus evolutivos. Em relação à funcionalidade, a IA pode referir-se a máquinas reativas, de memória limitada, cientes e autoconscientes. Na perspectiva da capacidade, a IA pode ser fraca, forte, geral, específica ou a super IA (LAGE, 2022, p. 50-52). Com a recente difusão do ChatGPT, pela OpenAI, a questão ganha novos contornos, dado o alto potencial disruptivo da tecnologia, ao alcance quase irrestrito de qualquer interessado.

Compreendidas as bases gerais da ODR, cumpre questionar se seria esse instrumento capaz de conferir maior efetividade ao controle externo. Não haveria empecilhos ao uso de plataformas voltadas à solução consensual de conflitos no âmbito público (indisponibilidade do interesse público)? A partir dessas indagações, vislumbra-se a possibilidade de conferir uma aplicação disruptiva dessa tecnologia aos processos de denúncia, de competência dessas Cortes.

Como é cediço, a Constituição da República de 1988 estabelece que qualquer cidadão é parte legítima para denunciar irregularidades ou ilegalidades perante os Tribunais de Contas (art. 74, § 2º, c/c 75). As respectivas Leis Orgânicas dessas Cortes estabelecem as diretrizes a serem observadas pelo cidadão e o

respectivo procedimento a ser desenvolvido no âmbito de cada Corte. Regra geral, essas normas foram estabelecidas num contexto de tramitação de processos físicos, a partir de uma lógica tradicional de fases a serem superadas sucessivamente e em que a movimentação processual e as decisões dependiam exclusivamente da manifestação de vontade humana.

Com a Lei nº 11.419/2006, que dispôs sobre a informatização do processo judicial, e em razão do caráter parcialmente “judicialiforme” dos processos de contas, algumas mudanças passaram a ser percebidas nos Tribunais de Contas, tais como a virtualização de autos de processo, a prática eletrônica de atos processuais e, mais recentemente, até mesmo o uso da inteligência artificial para subsidiar a prática de atos e decisões no âmbito do controle externo (GONZALEZ, 2023).

Considerando tal cenário, é possível propor o uso das plataformas de ODR no contexto dos Tribunais de Contas, em especial visando promover a solução de controvérsias decorrentes da propositura de denúncias em face dos agentes públicos fiscalizados. A possibilidade do uso da ODR no âmbito do setor público já fora demonstrada anteriormente (AMORIM; RODRIGUES, 2019), todavia agora o enfoque recai especificamente sobre a atuação dos Tribunais de Contas.

Atualmente, observa-se a inserção cada vez maior da consensualidade no Direito Público. Diversas leis passaram a permitir, no âmbito da Administração Pública, a negociação como forma de resolução de conflitos. A Lei de Improbidade Administrativa (LIA), Lei nº 8.429/1992, que restringia expressamente a possibilidade de transação, acordo ou conciliação em seu âmbito, passou a permitir a celebração de acordo de não persecução cível e a possibilidade de solução consensual no curso do processo (art. 17, §§ 1º e 10-A, da LIA, após a Lei nº 13.964/2019, e, atualmente, no art. 17-B, após a Lei nº 14.230/2021).

A Lei de Introdução às Normas do Direito Brasileiro (Lindb), Decreto-Lei nº 4.657/1942, após as alterações promovidas pela Lei nº 13.655/2018, regulamentada pelo Decreto nº 9.830/2019, passou a prever a possibilidade de a Administração Pública celebrar compromisso com particulares, para eliminar irregularidade, incerteza jurídica ou situação contenciosa na aplicação do direito público (art. 30).

Percebe-se a existência de um ambiente favorável à (i) implementação de soluções inovadoras baseadas em novas tecnologias para a resolução de conflitos no âmbito da Administração Pública, por meio de plataformas digitais; (ii) uma maior abertura para a resolução de conflitos entre o poder público e particulares a partir da consensualidade; (iii) um incentivo à participação do cidadão na fiscalização e no controle da Administração Pública.

Esta conjuntura justifica o desenvolvimento de pesquisas voltadas à delimitação das possibilidades jurídicas ao alcance dos agentes públicos, a fim de buscarem o desenvolvimento de soluções inovadoras. Isto tanto numa perspectiva teórica e dogmática, como também empírica, que identifique o atual estado da arte das inovações tecnológicas no âmbito do controle da Administração Pública, sem prejuízo de ser propositiva, ao estimular iniciativas que fomentem a evolução do controle externo. O estudo da realidade dos Tribunais de Contas será fundamental para avaliar como estas Cortes vêm desenvolvendo a transformação digital necessária ao ingresso na era do Indústria 4.0 e do Governo Digital.

Como destacam Gong, Yang e Shi, em estudo de caso da província chinesa de Zhejiang, para alcançar a transformação digital, as organizações precisam ter flexibilidade – capacidade de responder de forma adequada a um ambiente de mudanças –, que pode ser induzida pela via do uso das novas tecnologias ou de decisões políticas (2020). Imprescindível, portanto, avaliar em que estágio

tecnológico tais Cortes se encontram, bem como a existência de decisões no sentido de buscar o ingresso nessa nova era.

CONCLUSÕES

Os avanços da tecnologia alcançam todos os setores da sociedade; os Tribunais de Contas não estão imunes a tais mudanças. Uma postura receptiva e proativa é essencial para que estas Cortes preservem a sua relevância num mundo onde o digital se torna cada vez mais presente.

No contexto brasileiro, as bases para a transformação digital do poder público foram lançadas pela Lei do Governo Digital, marco regulatório importante e capaz de conferir limites, assegurar direitos aos usuários e induzir a inovação no ambiente público, inclusive nos Tribunais de Contas.

Há um ambiente propício, portanto, a mudanças que aliem o uso das novas tecnologias a uma maior participação da sociedade no controle social de viés consensual. Nesse contexto, a inteligência artificial aplicada às plataformas de *Online Dispute Resolution*, adaptadas às peculiaridades das Cortes de Contas, indica um caminho promissor a ser seguido, inclusive com experiências reais adotadas pelo Poder Judiciário em outros países, que poderiam ser replicadas no contexto do controle externo brasileiro.

Um ponto de partida palpável a aliar todos esses aspectos – controle social, consensualidade e novas tecnologias – corresponde à implementação de plataformas de *ODR* voltadas a solucionar as controvérsias decorrentes das denúncias apresentadas por cidadãos aos Tribunais de Contas do país, providência que, a um só tempo, poderia engajar e maximizar de forma expressiva a capacidade de fiscalização do controle externo, mudando sua forma de atuar, num sentido verdadeiramente disruptivo, cujo detalhamento impõe o necessário aprofundamento e desdobramento em trabalhos futuros.

REFERÊNCIAS

- AMORIM, Fernando Sérgio Tenório de; RODRIGUES, Ricardo Schneider. A resolução *online* de litígios (ODR) na administração pública: o uso da tecnologia como estímulo à transparência. *Direito, Estado e Sociedade*, n. 54, p. 171-204, jan./jun. 2019. Disponível em: <http://direitoestadosociedade.jur.puc-rio.br/media/Art7%20Ed54.pdf>. Acesso em: 24 jun. 2021.
- ARBIX, Daniel do Amaral. *Resolução Online de Controvérsias*. São Paulo: Intelecto, 2017.
- ATRICON. Nota Recomendatória Conjunta ATRICON/IRB/ABRACOM/CNPTC/AUDICON n° 2/2022. *Atricon*, Brasília, 13 out. 2022. Disponível em: <https://atrimon.org.br/wp-content/uploads/2022/10/Nota-Recomendatoria-Conjunta-no-02-2022-Lei-do-Governo-Digital-5.pdf>. Acesso em: 15 abr. 2023.
- CASTELLS, Manuel, 2010. Globalisation, Networking, Urbanisation: Reflections on the Spatial Dynamics of the Information Age. *Urban Studies*, v. 47, n. 13, 2010.
- CRT – Civil Resolution Tribunal. The CRT Process. *CRT*, 2023. Disponível em: <https://civilresolutionbc.ca/crt-process/>. Acesso em: 15 abr. 2023.
- DONEDA, Danilo. *Da Privacidade à Proteção de Dados Pessoais*. 2. ed. São Paulo: Revista dos Tribunais, 2019.
- DONEDA, Danilo. A proteção dos dados pessoais como um direito fundamental. *Espaço Jurídico*, Joaçaba, v. 12, n. 2, p. 91-108, jul./dez. 2011.
- GONG, Yiwei; YANG, Jun; SHI, Xiaojie. Towards a comprehensive understanding of digital transformation in government: Analysis of flexibility and enterprise architecture. *Government Information Quarterly*, n. 37, 2020. DOI: <https://doi.org/10.1016/j.giq.2020.101487>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0740624X19304174?via%3Dihub>. Acesso em: 23 jul. 2021.
- GONZALEZ, Mariana. TCU terá nova ferramenta de inteligência artificial capaz de ler e redigir despachos. *Jota*, Tecnologia GPT3, 22 mar. 2023. Disponível em: https://www.jota.info/legislativo/tcu-tera-nova-ferramenta-de-inteligencia-artificial-capaz-de-ler-e-redigir-despachos-22032023?utm_campaign=jota_info_ultimas_noticias_destaque_22032023&utm_medium=email&utm_source=RD+Station. Acesso em: 15 abr. 2023.

LAGE, Fernanda. *Manual de Inteligência Artificial no Direito Brasileiro*. 2. ed. São Paulo: Juspodivm, 2022.

LOSS, Juliana; ARBIX, Daniel. *Resolução Online de Disputas: casos brasileiros*. Rio de Janeiro: FGV Editora, 2022.

MALONE, Hugo; NUNES, Dierle. *Manual da Justiça Digital: compreendendo a Online Dispute Resolution e os Tribunais Online*.

MARQUES NETO, Floriano de Azevedo; FREITAS, Rafael Vêras de. Uber, WhatsApp, Netflix: os novos quadrantes da *publicatio* e da assimetria regulatória. In: FREITAS, Rafael Vêras de; RIBEIRO, Leonardo Coelho; FEIGELSON, Bruno (Coord.). *Regulação e Novas Tecnologias*. Belo Horizonte: Fórum, 2017. p. 17-48.

OLIVEIRA, Priscila. Tribunais de Contas investem na inteligência artificial para ampliar as ações de fiscalização. *Atricon*, Rádio Justiça, 12 jul. 2022. Disponível em: <https://atrimon.org.br/tribunais-de-contas-investem-na-inteligencia-artificial-para-ampliar-as-acoes-de-fiscalizacao/>. Acesso em: 15 abr. 2023.

PUTRIJANTI, Aju. Participation of Society In Decision-Making by Government in Industrial Revolution 4.0. *E3S Web of Conferences*, n. 125, 28 out. 2019. DOI: <https://doi.org/10.1051/e3sconf/201912502018>. Disponível em: https://www.e3s-conferences.org/articles/e3sconf/abs/2019/51/e3sconf_icenis2019_02018/e3sconf_icenis2019_02018.html. Acesso em: 22 jul. 2021.

RISTIANDY, Risky. Bureaucratic Disruption and Threats of Unemployment in the Industrial Revolution 4.0 Era. *Journal of Local Government Issues - LOGOS*, v. 3, n 1, mar. 2020. Doi: <https://doi.org/10.22219/logos.v3i1.10923>. Disponível em: <http://ejournal.umm.ac.id/index.php/LOGOS>. Acesso em: 23 jul. 2021.

RUSSELL, Stuart. *Inteligência artificial a nosso favor: como manter controle sobre a tecnologia*. Tradução Berilo Vargas. São Paulo: Companhia das Letras, 2021.

CIUDADES, TIC E INTELIGENCIA ARTIFICIAL

Álvaro Sánchez Bravo

Facultad de Derecho - Universidad de Sevilla. España

INTRODUCCIÓN. CIUDADES Y TIC

Todos los responsables políticos, al menos formalmente, se afanan en promocionar una ciudad atractiva y bella para el turista, cómoda y saludable para el habitante que vive en ella y, a ser posible, con un gran potencial económico. La realidad no siempre es así. A pesar de este interés -cualquiera que sea la ideología que impera- porque las ciudades sean referentes de sostenibilidad, de innovación, de creatividad, de alta tecnología y de compromiso con su población en crear vínculos de cuidado, no siempre las ciudades caminan por esos derroteros¹.

Las ciudades serán en los dos próximos siglos el modo principal de organización político-social, como lo han sido en los dos siglos anteriores las naciones, y antes los imperios. En efecto, vivimos un crecimiento sin precedentes de la urbanización y las ciudades se han convertido en los últimos decenios en un inexorable polo de atracción.

Ahora bien, mientras que, en Europa y América del Norte, la urbanización se desarrolló durante siglos, motivada por la industrialización y el aumento constante de la renta per cápita, en el mundo en desarrollo la se completará en el período de dos o tres generaciones.

¹ <https://fpablovi.org/tribuna/1697-ciudades-que-cuidan>

Ello ha llevado a que las ciudades, situadas en países de bajos ingresos, su desarrollo urbanístico viene acompañado de intolerables niveles de pobreza, desempleo e inseguridad. Y como ha señalado, Naciones Unidas, “la “ingobernabilidad” de las zonas urbanas responde más a fallas de gobernanza y de planificación urbana que al tamaño de las ciudades”²

Por su parte, a lo largo de los últimos tiempos, la industria tradicional, ha ido mutando hacia una economía basada en el conocimiento, en eso que se ha dado en llamar “nuevas tecnologías”.

Bajo la expresión “nuevas tecnologías” se esconden una pluralidad de procesos complejos que vienen a situarnos frente a nuevos retos, pero anclados en una cuestión plurisecularmente planteada: ¿qué papel puede jugar el desarrollo de la tecnología y de la ciencia, en la mejora de la vida de los seres humanos y de los grupos en que se integran? ³

La Declaración de Principios de la Cumbre Mundial de la Sociedad de la Información reunida en Ginebra ya manifestó, como respuesta a la cuestión planteada *“el compromiso común de construir una sociedad de la información centrada en la persona, incluyente y orientada a la desarrollo, en la que todos puedan crear, consultar y utilizar, y compartir la información y el conocimiento, para que las personas, las comunidades y los pueblos puedan desarrollar su pleno potencial en la promoción de su desarrollo sostenible y la mejora de su calidad de vida...”*⁴

² Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO), *Crear Ciudades más Verdes*, Roma, 2010, p. 12.

³ SANCHEZ BRAVO, A., *Internet y la sociedad europea de la información: implicaciones para los ciudadanos*, Publicaciones de la Universidad de Sevilla, 2001.

⁴ Cumbre Mundial de la Sociedad de la Información Ginebra 2003 - Túnez 2005. **Declaración de Principios. Construir la Sociedad de la Información: un desafío**

Porque de eso se trata, de no arrumbar al hombre como mero esclavo, sufridor o soporte de procesos tecnológicos, sino de hacer que sea el hombre el que encauce sus logros, corrija sus fracasos, y determine hacia dónde camina la técnica y la ciencia, con la vista puesta en una sociedad universal más digna, más libre y más justa.

Como punto de partida, es necesario considerar que pese a lo obvio que pudiera parecer, la tecnología no es, por sí misma, ni buena ni mala. Es la utilización que de la misma se haga la que determinara la naturaleza y extensión de sus beneficios o perjuicios.

Lo que, si no conviene obviar, *prima facie*, es que no es posible deslindar los aspectos sociales, de los aspectos técnicos, industriales y económicos que confluyen en la nueva sociedad tecnológica. En el futuro podrán existir diversos modelos de sociedad tecnológica, dependiendo de los diversos modelos que de industrialización se impongan, y que consecuentemente, deberán considerar diversos modelos y compromisos políticos para evitar la exclusión social y la creación de nuevas oportunidades para las personas y sectores más desfavorecidos.

Por otro lado, conviene tener presente como la revolución tecnocientífica está propiciando profundas transformaciones en el modo de concebir nuestras sociedades, su organización y estructuras. Efectivamente, los procesos tecnológicos están en constante evolución, sus capacidades aumentan día a día, su utilización se multiplica y diversifica sin cesar, y se imbrica en múltiples sectores a los que debe prestarse atención.

Desde la década de los 70, los progresos tecnológicos se han evidenciado con mayor énfasis en el campo de las TIC. En la hora presente, el progreso técnico nos permite tratar, almacenar y transmitir información, en cualquiera de sus formatos con

mundial para el nuevo milenio. Documento WSIS-03/GENEVA/DOCS/4-S, 12 de diciembre de 2003.

independencia de la distancia, tiempo y volumen. En este ámbito, Internet ha devenido el símbolo emblemático de ese nuevo proceso de globalización planetaria que mueve el engranaje de la consecución de una sociedad mundial de la información.

Ahora bien, esa globalización para que sea correctamente justipreciada requiere que no sea considerada de manera ideal, pues junto a innegables progresos y mejoras, ha puesto en evidencia fenómenos de agresión a los derechos y libertades de los ciudadanos.

Lo relevante en este punto sea considerar que la opción por una sociedad global del conocimiento, deba ser entendida como una *sociedad mundial que comparte el conocimiento*. Sólo de esta forma sería aceptable un movimiento globalizador que sirviera para unir a los pueblos, para caminar por sendas comunes, para resolver problemas universales, y no como lo que hoy se esconde bajo esa expresión sutil, pero a la vez terrible, que intenta imponer modelos de conducta, unidireccionalidad del pensamiento, y que busca eliminar a los diferentes, a los no productivos, a todos aquellos que no se someten al “conocimiento” del “imperio”.

No se olvide que la comunicación es un proceso social fundamental, una necesidad humana básica y el fundamento de toda organización social, siendo indispensable para la articulación de la sociedad de la información. Es por ello que, como ha señalado Naciones Unidas, *todas las personas en todas partes, deben poder participar y no debe excluirse a nadie de las ventajas que ofrece la sociedad de la información*.

Pero la participación en la información no debe basarse en actitudes paternalistas o “caritativas”, sino que debe operarse un salto cualitativa y cuantitativamente importante: “el paso de la solidaridad activa a la solidaridad pasiva”. Las nuevas tecnologías no serán más que fuegos de artificio, si no contribuyen a una sociedad más equitativa, más igualitaria, donde su potencial se

haga más accesible a todos, y contribuya a la mejora de los niveles de calidad de vida de los ciudadanos.

INTELIGENCIA ARTIFICIAL. ESTRATEGIAS ÉTICAS

Cada tres años se dispone de más información nueva que la creada en toda la historia de la humanidad. El único modo de gestionar esa información es mediante el uso de tecnologías digitales.

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 y 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.^{5 6}

⁵ 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.

⁶ 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:

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 refiere 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

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 XXI⁷.

Ahora bien, debe establecerse un enfoque sólido y conjunto, y abordar los retos que presenta la IA. La idea es triple: avanzar en el conocimiento científico; ahondar en el liderazgo tecnológico, y, sobre todo, garantizar que las nuevas tecnologías de IA estén al servicio de todos los ciudadanos, de una manera inclusiva y garantizando sus derechos.

Pero para ello, el punto de partida debía ser, dado el enorme impacto de la IA en la vida de las personas, generar confianza en su uso, y para ello, debe cimentarse en los valores y derechos

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.

⁷ SÁNCHEZ BRAVO, A., Derecho, Inteligencia Artificial y Nuevos Entornos Digitales, edic. a cargo de A. Sánchez Bravo, Punto Rojo Libros, Sevilla, 2020.

fundamentales⁸, como la dignidad humana y la protección de la intimidad.

La confianza es, por tanto, un requisito imprescindible y previo para garantizar que la IA esté garantizada y los ciudadanos confíen en ella.

Pero la IA presenta nuevos desafíos, ya que permite a las máquinas “aprender” y tomar decisiones y llevarlas a término sin acción humana. Ahora bien, las decisiones adaptadas “autónomamente” pueden dar como resultado datos incompletos, y por tanto, no fiables, que puedan ser manipulados por ciberataques, puedan ser datos sesgados o simplemente estar equivocados. Aplicar la tecnología de forma irreflexiva o poco cuidadosa generará resultados problemáticos, y ataques lesivos a los derechos de los ciudadanos, lo que supondrá su rechazo.

Por tanto, todas las acciones de IA deben ajustarse no sólo a la ley, sino también respetar unos principios éticos y garantizar que su implementación no genera daños involuntarios. En cada una de las fases de desarrollo de la IA deben garantizarse los derechos de los ciudadanos, y especialmente, la diversidad en cuanto al género, origen racial o étnico, religión o creencias, discapacidad y edad.

⁸ En el ámbito de la Unión Europea, vid. Carta de los Derechos Fundamentales de la Unión Europea, DOUE C 83. 30.03.2010. En el campo de la IA, y para abordar los retos y aprovechar al máximo las oportunidades, en abril de 2028 la Comisión Europea, inicia las actividades comunitarias con la publicación de una Estrategia europea (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. COM (2018) 237. Bruselas. 25.04.2018). Esta primera iniciativa europea destaca por colocar a la persona en el centro del desarrollo de la IA, adoptando un planteamiento triple para potenciar la capacidad tecnológica e industrial de la UE e impulsar la IA en todos los ámbitos económicos, prepararse para las transformaciones socioeconómicas y garantizar la aprobación de un marco ético y jurídico adecuado.

Debe quedar claro que la IA debe mejorar y apoyar la capacidad de las personas, pero no sustituirla.

Es por ello que se impone la elaboración de unas directrices éticas aplicables a desarrolladores, proveedores y usuarios de IA en el UE estableciendo unas condiciones de competencia ética en todos los Estados⁹.

La fiabilidad de la IA se apoya en tres componentes: que debe cumplirse a lo largo de todo el ciclo de vida del sistema. Estos tres componentes son imprescindibles cada uno de ellos, pero por sí mismos no son suficientes para garantizar una IA fiable. Estos componentes son:

a) la IA debe ser lícita; es decir, cumplir todas las leyes y reglamentos aplicables. Las Directrices no profundizan en este apartado, remitiéndose al Derecho primario y secundario de la UE y las normas y acuerdos internacionales aplicables.

b) la IA ha de ser ética. Deben desarrollarse, desplegarse y utilizarse los sistemas de inteligencia artificial respetando los principios éticos de respeto de la autonomía humana¹⁰, prevención

⁹ Directrices Éticas para una IA fiable.

<https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8cf-01aa75ed71a1/language-es/format-PDF/source-121796438>

¹⁰ Los derechos fundamentales en los que se apoya la UE van dirigidos a garantizar el respeto de la libertad y la autonomía de los seres humanos. Las personas que interactúen con sistemas de IA deben poder mantener una autonomía plena y efectiva sobre sí mismas y ser capaces de participar en el proceso democrático. Los sistemas de IA no deberían subordinar, coaccionar, engañar, manipular, condicionar o dirigir a los seres humanos de manera injustificada. En lugar de ello, los sistemas de IA deberían diseñarse de forma que aumenten, complementen y potencien las aptitudes cognitivas, sociales y culturales de las personas.

del daño¹¹, equidad¹² y explicabilidad¹³, reconociendo y desarrollando mecanismos de resolución de conflictos que puedan surgir entre estos principios¹⁴.

¹¹ Los sistemas de IA no deberían provocar daños (o agravar los existentes) ni perjudicar de cualquier otro modo a los seres humanos. Esto conlleva la protección de la dignidad humana, así como de la integridad física y mental. Todos los sistemas y entornos de IA en los que operan estos deben ser seguros. También deberán ser robustos desde el punto de vista técnico, y debería garantizarse que no puedan destinarse a usos malintencionados. Las personas vulnerables deberían recibir mayor atención y participar en el desarrollo y despliegue de los sistemas de IA. Se deberá prestar también una atención particular a las situaciones en las que los sistemas de IA puedan provocar efectos adversos (o agravar los existentes) debido a asimetrías de poder o de información, por ejemplo, entre empresarios y trabajadores, entre empresas y consumidores o entre gobiernos y ciudadanos. La prevención del daño implica asimismo tener en cuenta el entorno natural y a todos los seres vivos. *Íbidem*.

¹² El desarrollo, despliegue y utilización de sistemas de IA debe ser equitativo. Pese a que reconocemos que existen muchas interpretaciones diferentes de la equidad, creemos que esta tiene tanto una dimensión sustantiva como procedimental. La dimensión sustantiva implica un compromiso de: garantizar una distribución justa e igualitaria de los beneficios y costes, y asegurar que las personas y grupos no sufran sesgos injustos, discriminación ni estigmatización. Si se pueden evitar los sesgos injustos, los sistemas de IA podrían incluso aumentar la equidad social. También se debería fomentar la igualdad de oportunidades en términos de acceso a la educación, los bienes los servicios y la tecnología. Además, el uso de sistemas de IA no debería conducir jamás a que se engañe a los usuarios (finales) ni se limite su libertad de elección. Asimismo, la equidad implica que los profesionales de la IA deberían respetar el principio de proporcionalidad entre medios y fines, y estudiar cuidadosamente cómo alcanzar un equilibrio entre los diferentes intereses y objetivos contrapuestos. La dimensión procedimental de la equidad conlleva la capacidad de oponerse a las decisiones adoptadas por los sistemas de IA y por las personas que los manejan, así como de tratar de obtener compensaciones adecuadas frente a ellas³². Con este fin, se debe poder identificar a la entidad responsable de la decisión y explicar los procesos de adopción de decisiones. *Ídem*, pp. 15-16.

¹³ Esto significa que los procesos han de ser transparentes, que es preciso comunicar abiertamente las capacidades y la finalidad de los sistemas de IA y que las decisiones deben poder explicarse —en la medida de lo posible— a las partes que se vean afectadas por ellas de manera directa o indirecta. Sin esta información, no es posible impugnar adecuadamente una decisión. No siempre resulta posible explicar por qué un modelo ha generado un resultado o una decisión particular (ni qué combinación de factores contribuyeron a ello). Esos

Deberá prestarse especial atención a las situaciones que afecten a los grupos de población más vulnerables¹⁵, así como a aquellas situaciones donde se de una manifiesta situación de desequilibrio de poder o información.

c) la IA debe ser robusta; es decir, sólida desde el punto de vista técnico (adecuada en un contexto determinado, como el ámbito de aplicación o la fase de ciclo de vida) como social (que garantice que

casos, que se denominan algoritmos de «caja negra», requieren especial atención. En tales circunstancias, puede ser necesario adoptar otras medidas relacionadas con la explicabilidad (por ejemplo, la trazabilidad, la auditabilidad y la comunicación transparente sobre las prestaciones del sistema), siempre y cuando el sistema en su conjunto respete los derechos fundamentales. El grado de necesidad de explicabilidad depende en gran medida del contexto y la gravedad de las consecuencias derivadas de un resultado erróneo o inadecuado. Idem, p.16

¹⁴Cabe la posibilidad de que surjan tensiones entre los principios anteriores, y no existe una solución establecida para resolverlas. En consonancia con el compromiso fundamental de la UE con la participación democrática, el respeto de las garantías procesales y la participación abierta en la esfera política, deberían establecerse métodos que posibiliten un debate responsable sobre dichas tensiones. Además, los beneficios globales de los sistemas de IA deberían ser sustancialmente superiores a los riesgos individuales previsible.

¹⁵ Debido a su heterogeneidad, no existe una definición generalmente aceptada ni que cuente con un consenso amplio del concepto de «personas vulnerables». Lo que se considera una persona o grupo vulnerable suele depender del contexto. Los sucesos vitales de carácter temporal (como la infancia o la enfermedad), los factores de mercado (como la asimetría de información o el poder de mercado), los factores económicos (como la pobreza), los vinculados a la identidad de las personas (como el género, la religión o la cultura) y otros pueden desempeñar un papel en ese sentido. La Carta de los Derechos Fundamentales de la Unión Europea recoge en su artículo 21, relativo a la no discriminación, los motivos de discriminación siguientes, que pueden servir como punto de referencia, entre otros: el sexo, la raza, el color, los orígenes étnicos o sociales, las características genéticas, la lengua, la religión o las convicciones, las opiniones políticas o de cualquier otro tipo, la pertenencia a una minoría nacional, el patrimonio, el nacimiento, la discapacidad, la edad o la orientación sexual. En las disposiciones de otras leyes se abordan los derechos de determinados grupos, además de los enumerados anteriormente. Este tipo de listas nunca pueden ser exhaustivas, y pueden cambiar a lo largo del tiempo. Un grupo vulnerable es un grupo de personas que comparten una o varias características de vulnerabilidad. Idem, p.51.

el sistema de IA tenga adecuadamente el contexto y el entorno en que opera)¹⁶.

Los principios indicados deben traducirse en requisitos que permitan de manera efectiva conseguir una IA fiable. Los mismos deben ser aplicables a todas las partes que intervienen en algún momento en el ciclo de vida de sistemas de IA: desarrolladores, responsables de despliegue y usuarios finales, así como a la sociedad en su conjunto¹⁷.

Así, cada una de esas partes desempeñarán diversos papeles para garantizar el cumplimiento de los requisitos: a) los desarrolladores deben incluir y aplicar los principios en la fase de diseño y desarrollo; b) los responsables del despliegue deben asegurarse de que los sistemas utilizados y los productos y servicios ofertados cumplen los requisitos imprescindibles; y, c) los usuarios y la sociedad deben estar informados sobre los requisitos exigidos a los anteriores, y disponer de mecanismos para exigir su cumplimiento.

Siete son los requisitos esenciales que debe cumplir y respetar cualquier aplicación de IA para ser fiable:

- Intervención y supervisión humana. Los sistemas de IA deben ayudar a las personas a escoger mejor y con más conocimiento de sus opciones, en función de sus objetivos. Con ello se garantiza que un sistema de IA no menoscaba la autonomía humana, ni causa efectos adversos. Por ello deberán incrementarse medidas

¹⁶ Ídem, pp. 17-30.

¹⁷ El término «desarrolladores» hace referencia a las personas dedicadas a la investigación, el diseño o el desarrollo de sistemas de IA. Por «responsables del despliegue» se entienden las organizaciones públicas o privadas que utilizan sistemas de IA en sus procesos internos y para ofrecer productos y servicios a otros agentes. Los «usuarios finales» son aquellos que interactúan con el sistema de IA, ya sea de forma directa o indirecta. Por último, la «sociedad en su conjunto» engloba el resto de agentes, personas y entidades afectados de manera directa o indirecta por los sistemas de IA. *Ibidem*.

estrictas de control, incluida la adaptabilidad, la exactitud y la explicabilidad. Al respecto las autoridades públicas deben establecer funciones de supervisión mediante sistemas de gobernanza¹⁸, que determinen, entre otras cosas, que cuanto más difícil sea para un ciudadano controlar un sistema de IA, más estricta será la gobernanza.

- Solidez y seguridad técnicas. Los algoritmos deben ser seguros, fiables y sólidos para resolver los errores o incoherencias durante todo el ciclo vital del sistema de IA, solucionando adecuadamente los errores que se produzcan. Al respecto, deberán elaborarse mecanismos de seguridad, incluyendo la seguridad desde el diseño, para verificar que son objetivamente verificables en cada fase del proceso, teniendo muy en cuenta la seguridad física y psicológica de los usuarios. Igualmente, debe garantizarse la minimización, y cuando sea posible, la reversibilidad de consecuencias no deseadas o errores causados por el sistema¹⁹.

- Privacidad y gestión de datos. La protección de los datos personales debe garantizarse en todas las fases del ciclo vital del sistema de IA. Esto incluye los datos inicialmente suministrados por el usuario, así como la información generada sobre el usuario en el

¹⁸ Entre los sistemas de gobernanza pueden indicarse, Human-in-the-loop (HITL) se refiere a la intervención humana en cada ciclo de decisión del sistema, lo que en muchos casos no es posible ni deseable. Human-on-the-loop (HOTL) se refiere a la capacidad de la intervención humana durante el ciclo de diseño del sistema y a la supervisión del funcionamiento del sistema. Human-in-command (HIC) se refiere a la capacidad de supervisar la actividad global del sistema de IA (incluido su impacto más amplio económico, social, jurídico y ético) y a la capacidad de decidir cuándo y cómo utilizar el sistema en cada situación determinada. Esto puede incluir la decisión de no utilizar un sistema de IA en una situación concreta, establecer niveles de discreción humana durante el uso del sistema o garantizar la capacidad de imponerse a una decisión tomada por el sistema. COMUNICACIÓN DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO EUROPEO, AL CONSEJO, AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO Y AL COMITÉ DE LAS REGIONES. Generar confianza en la inteligencia artificial centrada en el ser humano, cit. p.5.

¹⁹ *Ibidem*.

contexto de su utilización del sistema. Como sabemos, los registros digitales del comportamiento humano, pueden producir no sólo el suministro de preferencias al usuario, sino también que la IA infiera su orientación sexual, edad, género u opiniones políticas o religiosas. Es por ello, que es necesario garantizar que los datos recogidos no se utilizarán para discriminar a los usuarios de forma injusta o ilegal.

Será igualmente necesario garantizar la calidad e integridad de los datos, así como el acceso a los mismos²⁰.

- Transparencia. Debe garantizarse la trazabilidad de los sistemas de IA, registrar y documentar las decisiones tomadas por los sistemas como la totalidad del proceso que dio lugar a la decisión. Igualmente debe aportarse la explicabilidad del proceso

²⁰ Directrices Éticas para una IA fiable, cit., p.22.

Reglamento (UE) 2016/679 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. DOUE L 119. 04.05.2016.

DIRECTIVA (UE) 2016/680 DEL PARLAMENTO EUROPEO Y DEL CONSEJO, de 27 de abril de 2016, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por parte de las autoridades competentes para fines de prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de sanciones penales, y a la libre circulación de dichos datos y por la que se deroga la Decisión Marco 2008/977/JAI del Consejo. DOUE L 119. 04.05.2016.

Vid. Sánchez Bravo, A. "Nuevo marco europeo de protección de datos personales", en la obra colectiva, edic. a cargo de Álvaro Sánchez Bravo, Derechos humanos y protección de datos personales en el Siglo XXI:homenaje a Cinta Castillo Jiménez, Punto Rojo Libros, Sevilla, 2014, pp. 255-288; y, "Hacia un nuevo marco europeo de protección de datos personales: empoderamiento de los ciudadanos en la sociedad tecnológica", en la obra colectiva, coord. a cargo de Yarina Amoroso Fernández, Sociocibernética e Infoética: contribución a una nueva cultura y praxis jurídica, Editorial UNIJURIS, La Habana, 2015. pp. 108-133. REGLAMENTO (UE) 2018/1807 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 14 de noviembre de 2018 relativo a un marco para la libre circulación de datos no personales en la Unión Europea. DOUE L 303. 28.11.2018.

algorítmico de tomas de decisiones, pero adaptada al nivel de las personas afectadas.

Deberá comunicarse adecuadamente las capacidades y limitaciones del sistema de IA a las diferentes partes interesadas afectadas conforme al caso de que se trate²¹.

- Diversidad, no discriminación y equidad. Los conjuntos de datos que alimentan los sistemas de inteligencia artificial pueden verse afectados por la inclusión de sesgos históricos involuntarios, por no ser completos o por no haberse implementado modelos de gobernanza adecuados.

Puede ayudar a resolver estos problemas establecer equipos de diseño diversificados y crear mecanismos que garanticen la participación, en particular de los ciudadanos, en el desarrollo de la IA. Los sistemas de IA deberían tener en cuenta toda la gama de capacidades, habilidades y necesidades humanas y garantizar la accesibilidad mediante un enfoque de diseño universal para tratar de lograr la igualdad de acceso para las personas con discapacidades²².

- Bienestar social y medio ambiental. Los sistemas de inteligencia artificial deben utilizarse para mejorar el cambio social positivo y aumentar la sostenibilidad y la responsabilidad ecológicas.

- Rendición de cuentas. Deben implantarse mecanismos que garanticen la responsabilidad y la rendición de cuentas de los sistemas de inteligencia artificial y de sus resultados.

²¹ . COMUNICACIÓN DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO EUROPEO, AL CONSEJO, AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO Y AL COMITÉ DE LAS REGIONES. Generar confianza en la inteligencia artificial centrada en el ser humano, cit. p.6.

²² Ibidem.

El planteamiento en materia de inteligencia artificial (IA), fundado en la confianza y la excelencia, ofrecerá a los ciudadanos la seguridad necesaria para adoptar estas tecnologías, animando al mismo tiempo a las empresas a desarrollarlas²³.

CIUDADES E INTELIGENCIA ARTIFICIAL

Las *Smart Cities* o ciudades inteligentes son, como indica la Comisión Europea²⁴, lugar que integra los sistemas físicos, digitales y humanos en las redes y servicios tradicionales para utilizar mejor los recursos energéticos y reducir las emisiones en beneficio de los ciudadanos y las empresas.

La definición de «ciudad inteligente» va más allá del uso de las tecnologías digitales: también incluye edificios más eficientes desde el punto de vista energético, fuentes de energía renovables integradas, sistemas de calefacción y refrigeración sostenibles, redes de transporte urbano más inteligentes, la mejora del suministro de agua y mejores instalaciones de eliminación de residuos para hacer frente a los retos económicos, sociales y medioambientales de la ciudad. Las ciudades inteligentes dependen de un compromiso político y de un compromiso ciudadano amplio e integrador para ofrecer soluciones sostenibles e inclusivas para que las ciudades sean más resilientes.

Muchos modelos se han ensayado, y se proponen, respecto a las dimensiones básicas mínimas que deben tener estos modelos de ciudades, como adaptación de las urbes tradicionales a las TIC. Como elementos básicos, pueden considerarse los siguientes:

²³ https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/excellence-trust-artificial-intelligence_es#%C3%BAltimas-novedades

²⁴ Comisión Europea. Dirección General de Energía, *En el punto de mira: Energía y ciudades inteligentes*, Bruselas. 13 de junio de 2022.

https://commission.europa.eu/news/focus-energy-and-smart-cities-2022-07-13_es

a) Generación distribuida. Consiste en que las ciudades inteligentes cuenten con generación eléctrica repartida por el territorio, a través de un abastecimiento individualizado (micro-generación), no centralizado.

b) Smart Grids. Se conoce como Smart Grids a las redes inteligentes interconectadas, las cuales poseen una circulación bidireccional de datos entre el service center (o centro de control) y el usuario.

Smart Metering. Se trata de la medición inteligente de los datos de gasto energético de cada usuario, a través de telecontadores donde se realizan las lecturas a distancia y a tiempo real.

Smart Buildings. Los edificios de la Smart City se convierten en inteligentes como modelo de eficiencia. Edificios domóticos que respetan el medioambiente y que poseen sistemas de producción de energía integrados.

Smart Sensors. Los sensores inteligentes tienen la función de recopilar todos los datos necesarios para hacer de la ciudad una Smart City. Son parte fundamental para mantener la ciudad conectada e informada, y hacer que cada subsistema cumpla su función.

eMobility. Implantación del vehículo eléctrico y los respectivos puestos de recarga públicos y privados.

Tecnologías de la información y la comunicación (TIC). Son las tecnologías de la información que ayudan a la hora de controlar los diferentes subsistemas que componen la Smart City. Mediante ellos los ciudadanos y las entidades administrativas pueden participar activamente en el control de la ciudad.

Smart Citizen. Los ciudadanos son sin duda la parte fundamental de una Smart City, ya que sin su participación activa no es posible poder llevar a cabo estas iniciativas²⁵.

Ahora bien, como veníamos reiterando a lo largo de esta modera reflexión, la participación ciudadana debe ser el eje en el que pivoten las nuevas estructuras que se pretenden desarrollar o implementar, bajo el paraguas de la inteligencia artificial, en nuestras ciudades.

Sin ánimo de ser exhaustivos, debemos considerar los pros y los contras que estas tecnologías pueden acarrear en esta titánica tarea.

Así, uno de los mayores desafíos es la eliminación de empleos humanos debido a la automatización. A medida que las empresas encuentran formas de automatizar tareas, como la fabricación o la atención al cliente, es probable que muchas personas pierdan sus trabajos, generando de nuevo, como en la época de la Revolución Industrial en Europa, levas de personas desempleadas, o en empleos precarios, que serían la base para el mantenimiento de guetos territoriales y sociales²⁶.

Otro desafío, como indicamos, lo constituye el riesgo de atentados a los derechos humanos vinculados a las nuevas tecnologías, especialmente la privacidad y la seguridad de los datos. La IA necesita, se alimenta, de grandes cantidades de datos y, si estos datos no se manejan adecuadamente, pueden ser la base para

²⁵ <https://www.fundacionendesa.org/es/educacion/endesa-educa/recursos/smart-city>

²⁶ Sánchez Bravo, A., Covid e impactos socio-laborales: reflexiones desde la filosofía del derecho, en la obra colectiva edic. a cargo de Francisco Javier Calvo Gallego, Macarena Hernández-Bejarano y Miguel Rodríguez-Piñero Royo, *La revolución de las formas de empleo en el siglo XXI*, Laborum, 2021, pp. 21-36.

sistemas tecnológicos que dañen a las personas. Además, existe el riesgo de que los sistemas de IA tomen decisiones sesgadas o discriminatorias, lo que podría tener consecuencias negativas para grupos enteros de personas, convirtiendo la ciudad en un territorio hostil y segregado, basado en la decisión de un algoritmo²⁷.

Es importante que los gobiernos, las empresas y los ciudadanos trabajen juntos para educar al público sobre los beneficios y riesgos de la IA para que puedan tomar decisiones informadas sobre su uso. Especialmente preocupante es la situación respecto a las personas denominadas de la “tercera edad”, que ha generado un fenómeno conocido como “edadismo”²⁸.

Es palpable, la preocupación de que, si no se controlan, las tecnologías de IA puedan perpetuar la discriminación por edad existente en la sociedad y socavar la calidad de la atención médica y social que reciben las personas mayores. Los datos utilizados por la IA pueden no ser representativos de las personas mayores o estar sesgados por estereotipos, prejuicios o discriminación por motivos de edad del pasado. Las suposiciones erróneas sobre cómo las personas mayores desean vivir o interactuar con la tecnología en su vida diaria también pueden limitar el diseño y el alcance de estas tecnologías, y la forma en que se utilizan las tecnologías de IA puede reducir el contacto intergeneracional o profundizar las barreras existentes para el acceso digital.

²⁷ Sánchez Bravo, A., Inteligencia artificial, control y nuevos marcos normativos en la Unión Europea, en la obra colectiva edic. a cargo de Fernando Higinio Llano Alonso y Joaquín Garrido Martín, *Inteligencia Artificial y Derecho. El jurista ante los retos de la era digital*, Aranzadi, 2021, pp. 307-330.

²⁸ World Health Organization, Ageism in artificial intelligence for health: WHO policy brief., 2022.

[https://www.who.int/news/item/09-02-2022-ensuring-artificial-intelligence-\(ai\)-technologies-for-health-benefit-older-people](https://www.who.int/news/item/09-02-2022-ensuring-artificial-intelligence-(ai)-technologies-for-health-benefit-older-people)

Desde el punto de vista de las relaciones sociales, es reiterativo el reseñar el impacto de las TIC, y la IA. En algunos casos, ha mejorado la forma en que nos comunicamos y conectamos con los demás, mientras que en otros casos ha generado preocupación por el riesgo de aislamiento social y la falta de habilidades para interactuar con personas reales.

Por otro lado, estas mismas herramientas pueden llevarnos a una burbuja de información, donde solo se nos muestra contenido similar a nuestras preferencias, reforzando un pensamiento único, que nos aparte y excluya de la pluralidad y sana discrepancia de una sociedad democráticamente viva y plural²⁹.

En cuanto a las relaciones interpersonales, la IA ha demostrado ser efectiva en la identificación de emociones y expresiones faciales, lo que puede facilitar la comunicación y la empatía en algunos contextos. Sin embargo, también existe la preocupación de que el uso excesivo de tecnologías pueda hacer que perdamos la capacidad de leer y comprender las señales no verbales de las personas en situaciones cara a cara³⁰.

Por último, y en el ámbito económico, la IA manifiesta para transformar la forma en que se llevan a cabo las operaciones económicas y para mejorar la eficiencia y la eficacia del control económico.

En primer lugar, la IA puede utilizarse para analizar grandes cantidades de datos económicos, lo que puede ayudar a los responsables políticos a identificar patrones y tendencias

²⁹ Sánchez Bravo, A., Estado de derecho, infodemia y libertad de expresión en la Unión Europea, en la obra colectiva edic. a cargo de Álvaro Sánchez Bravo, *Sensibilidad, sociología y derecho: Libro homenaje al Prof. Dr. José Alcebiades de Oliveira Junior*, Alma Mater, 2021, pp. 23-65.

³⁰ Scientific American, *Especial Inteligencia Artificial*, Investigación y Ciencia, abril 2014.

importantes en la economía. Contribuyendo a prever eventos económicos importantes, como recesiones o aumentos en los precios de los productos básicos, lo que puede ayudar a los responsables políticos a tomar decisiones informadas para mitigar su impacto³¹.

Igualmente. para mejorar la gestión del riesgo financiero y para detectar actividades fraudulentas en las operaciones financieras. Por ejemplo, los algoritmos de aprendizaje automático pueden analizar grandes cantidades de datos financieros para identificar patrones de comportamientos sospechosos y alertar a los responsables de la regulación financiera para que tomen medidas preventivas.

Además, la IA puede ser utilizada para mejorar la eficiencia y la eficacia de las operaciones económicas, como la gestión de la cadena de suministro y la optimización de la producción. Esto puede ayudar a reducir los costos y mejorar la calidad de los productos y servicios, lo que puede mejorar la competitividad de las empresas y la economía en genera³²l.

En resumen, la IA tiene un gran potencial para mejorar el control económico y la eficiencia de las operaciones económicas. Sin embargo, también es importante tener en cuenta los riesgos potenciales asociados con el uso de la IA, como la falta de transparencia y la posible pérdida de empleo debido a la automatización. Por lo tanto, es importante que los responsables políticos y los líderes empresariales consideren cuidadosamente tanto los beneficios como los riesgos potenciales al implementar la IA en sus operaciones económicas.

³¹ https://www.eca.europa.eu/es/news/NEWS2303_20

³² García Rodríguez, Manuel J., Tecnologías digitales para el control de la contratación pública, en Auditoría Pública, Nº 79 (2022) .

UNA ULTIMA REFLEXIÓN CON EL PAPA FRANCISCO

La mitad de las personas viven en ciudades, y eso “tiene un fuerte impacto en la mentalidad, la cultura, los estilos de vida, las relaciones interpersonales y la religiosidad de las personas”. A mayor abundamiento, el Santo Padre manifiesta como a las ciudades, son “espacios magníficos de libertad y de realización humana”, pero también “espacios terribles de deshumanización y de infelicidad”³³.

“Parece precisamente que cada ciudad, incluso la que se muestra más floreciente y ordenada, tenga la capacidad de generar dentro de sí una oscura «anti-ciudad». Parece que junto a los ciudadanos también existen los no-ciudadanos: personas invisibles, pobres de recursos y calor humano, que habitan en «no-lugares», que viven de las «no-relaciones». Se trata de personas a las que nadie les dirige una mirada, una atención, un interés. No sólo son los «anónimos», son los «anti-hombres». Y esto es terrible³⁴.

Pero ante estos tristes escenarios, debemos recordar siempre que Dios no ha abandonado la ciudad; Él vive en la ciudad...es posible encontrar a Dios en el corazón de la ciudad. Esto es muy hermoso. Sí, Dios sigue estando presente también en nuestras ciudades, tan frenéticas y distraídas. Por eso es necesario no abandonarse jamás al pesimismo y al derrotismo, sino tener una mirada de fe sobre la

³³Papa Francisco: “Dios está presente inclusive en nuestras ciudades frenéticas y distraídas” <https://www.aciprensa.com/noticias/papa-francisco-dios-esta-presente-inclusive-en-nuestras-ciudades-freneticas-y-distraidas-10147>

³⁴ DISCURSO DEL SANTO PADRE FRANCISCO A LOS PARTICIPANTES EN LA PLENARIA DEL CONSEJO PONTIFICIO PARA LOS LAICOS. Sala Clementina. Sábado 7 de febrero de 2015.

https://www.vatican.va/content/francesco/es/speeches/2015/february/documents/papa-francesco_20150207_pontificio-consiglio-laici.html

ciudad, una mirada contemplativa «que descubra al Dios que habita en sus hogares, en sus calles, en sus plazas»³⁵.

Reitera el Papa Francisco lo que es santo y seña de su magisterio, como las ciudades deben ser un lugar de encuentro, de reconciliación, con especial atención a los más desfavorecidos y abandonados. Pero no sólo desde el punto de vista material, sino también del social y desde la dimensión moral.

Nuestras sociedades que caminan tan rápido, donde el hoy ya es ayer, y donde las ciudades parecen fagocitar a sus habitantes, necesitan de un nuevo pacto ciudadano que restañe las heridas de la segregación, el clasismo, la indiferencia y la soledad.

Como hemos apuntado a lo largo del texto, la inteligencia artificial puede contribuir a esta capital reestructuración. Pero, para ello, apelábamos, a la necesidad de establecer potentes cortafuegos que hicieran de la tecnología un instrumento, de nuevo, de discriminación. Cuando no, de eliminación de un humanismo imprescindible, que nos subyuga a la “infalibilidad” de la máquina, frente a la riqueza, variedad y solidaridad de la inteligencia natural de las criaturas.

Y es, en este punto, donde resuenan, de nuevo, con voz clara y contundente, las palabras del Santo Padre Francisco.

“La tecnología es de gran ayuda para la humanidad. Pensemos en los innumerables avances en los campos de la medicina, la ingeniería y las comunicaciones (cf. Encíclica *Laudato si'*, 102). Y al mismo tiempo que reconocemos los beneficios de la ciencia y la tecnología, vemos en ellas una prueba de la creatividad del ser humano y también de la nobleza de su vocación a participar responsablemente en la acción creadora de Dios (cf. *ibíd.*, 131).

³⁵ *Ibidem*.

Desde esta perspectiva, creo que el desarrollo de la inteligencia artificial y del aprendizaje automático tiene el potencial de aportar una contribución beneficiosa al futuro de la humanidad, no podemos descartarlo”³⁶.

No obstante, Francisco es muy consciente de los riesgos y retos a los que estas tecnologías nos pueden arrastrar, y por ello: “estoy seguro de que este potencial sólo se hará realidad si existe una voluntad coherente por parte de quienes desarrollan las tecnologías para actuar de forma ética y responsable. Me anima el compromiso de tantas personas que trabajan en estos campos para garantizar que la tecnología se centre en el ser humano, se funde en bases éticas durante el diseño del proyecto y tenga por finalidad el bien”.

Ante la gran pluralidad presente en el mundo actual, se impone “un diálogo integrador, en el que las personas busquen juntas la verdad, puede propiciar un verdadero consenso; y esto puede ocurrir si compartimos la convicción de que «en la realidad misma del ser humano y de la sociedad [...] hay una serie de estructuras básicas que sostienen su desarrollo y su supervivencia» (Enc. *Fratelli tutti*, 212). El valor fundamental que debemos reconocer y promover es el de la *dignidad de la persona humana* (cf. *ibid.*, 213).”³⁷

³⁶ Sánchez Bravo, A., Ecología integral y doctrina social de la Iglesia, en la obra colectiva edic. a cargo de Álvaro Sánchez Bravo, *Semper Sapiens: Libro homenaje al Prof. Dr. Felipe Rotondo Tornarí*, Alma Mater, 2022, pp. 19-37; y, Francisco y Evangelio socioambiental, en la obra colectiva edic. a cargo de Débora Ferrazzo, Ivone Fernandes Morcilo lixa y Lucas Machado Fagundes, *Teoría crítica e humanismo jurídico. O pensamento de Antonio Carlos Wolkmer*, Centro de Estudios Jurídicos y Sociales Mispát, 2022, pp. 747-767.

³⁷ DISCURSO DEL SANTO PADRE FRANCISCO A LOS PARTICIPANTES EN LOS "MINERVA DIALOGUES" ORGANIZADO POR EL DICASTERIO PARA LA CULTURA Y LA EDUCACIÓN. Sala del Consistorio. Lunes, 27 de marzo de 2023. <https://www.vatican.va/content/francesco/es/speeches/2023/march/documents/20230327-minerva-dialogues.html>

De esta manera, muestra su preocupación por “que los datos disponibles hasta ahora parezcan sugerir que las tecnologías digitales han servido para aumentar las desigualdades en el mundo. No sólo las diferencias de riqueza material, que son importantes, sino también las diferencias de acceso a la influencia política y social. Nos preguntamos: ¿son nuestras instituciones nacionales e internacionales capaces de exigir responsabilidades a las empresas tecnológicas por el impacto social y cultural de sus productos? ¿Existe el riesgo de que el aumento de la desigualdad socave nuestro sentido de la solidaridad humana y social? ¿Podríamos perder nuestro sentido de destino compartido? En realidad, nuestro objetivo es que el crecimiento de la innovación científica y tecnológica vaya acompañado de una mayor igualdad e inclusión social”³⁸.

No se trata ni de temer sin fundamento, ni aceptar asépticamente los avances tecnológicos, sino tener claro, hoy y siempre, qué en el centro de los procesos históricos, sociales, culturales y tecnológicos, la persona sea el alfa y el omega, la inexcusable referencia.

Porque como señala el Santo Padre, y concluimos, *“El concepto de dignidad humana —este es el núcleo— exige que reconozcamos y respetemos el hecho de que el valor fundamental de una persona no puede medirse con un conjunto de datos. En los procesos de toma de decisiones sociales y económicas, debemos ser cautos a la hora de confiar juicios a algoritmos que procesan datos recogidos, a menudo subrepticamente, sobre las personas y sus características y comportamientos pasados. Esos datos pueden estar contaminados por prejuicios sociales e ideas preconcebidas. Sobre todo, porque el comportamiento pasado de un individuo no debe utilizarse para negarle la oportunidad de cambiar, crecer y*

³⁸ *Ibidem.*

contribuir a la sociedad. No podemos permitir que los algoritmos limiten o condicionen el respeto a la dignidad humana, ni que excluyan la compasión, la misericordia, el perdón y, sobre todo, la apertura a la esperanza de cambio en el individuo”.

ISBN 979-83-96042-25-4



9 798396 042254 >

am Alma mater