



Análisis de la propuesta de reglamento sobre los principios éticos para el desarrollo, el despliegue y el uso de la inteligencia artificial, la robótica y las tecnologías conexas

ANALYSIS OF THE PROPOSAL FOR A REGULATION ON ETHICAL PRINCIPLES FOR THE DEVELOPMENT, DEPLOYMENT AND USE OF ARTIFICIAL INTELLIGENCE, ROBOTICS AND RELATED TECHNOLOGIES

Guillermo Lazcoz Moratinos*

G.I. Cátedra de Derecho y Genoma Humano

Universidad del País Vasco (UPV/EHU)

guillermo.lazcoz@ehu.eus 0000-0001-6567-045X

Recibido: 02 de noviembre 2020 | Aceptado: 21 de diciembre 2020

RESUMEN

El pasado 20 de octubre de 2020, el Parlamento Europeo aprobó una Resolución (2020/2012(INL)) con recomendaciones destinadas a la Comisión sobre inteligencia artificial, robótica y tecnologías conexas, entre las cuales, se recoge una propuesta legislativa para la tramitación de un Reglamento sobre los principios éticos para el desarrollo, el despliegue y el uso de dichas tecnologías. El contenido de esta propuesta deriva, sin lugar a duda, de la visión regulatoria que la Comisión Europea ha mantenido en documentos como el Libro Blanco sobre Inteligencia Artificial (COM(2020) 65 final) o las

PALABRAS CLAVE

Inteligencia artificial
Unión Europea
Algoritmos
Discriminación
Supervisión humana

* **Financiación y agradecimientos:** El texto ha sido elaborado gracias a la financiación de la ayuda FPU 16/06314 del Ministerio de Universidades y del Departamento de Educación del Gobierno vasco para apoyar las actividades de Grupos de Investigación del Sistema Universitario Vasco (IT 1066-16), y se ha desarrollado en el marco del proyecto de investigación RTI2018-099306-B-100 Ciberseguridad y Ciberdelitos, integrado en el Programa Estatal de Proyectos de I+D+i Retos de Investigación 2018”, cofinanciado por el Ministerio de Ciencia, Innovación y Universidades, la Agencia Estatal de Investigación y el Fondo Europeo de Desarrollo Regional (FEDER - “Una manera de hacer Europa”).

Quiero agradecer a Daniel Jove Villares (Investigador FPU de la Universidade da Coruña, UDC) sus comentarios, correcciones e interés en esta investigación. Es siempre un placer intercambiar penas y glorias de este camino académico que discurrimos en paralelo.

Directrices éticas para una IA fiable elaboradas por el Grupo independiente de expertos de alto nivel. Ante este nuevo horizonte normativo, se hace más necesario que nunca abordar una crítica constructiva de la propuesta, destacando la necesidad o no de reformular su marcado carácter de *soft law* a pesar de ubicarse en una fuente normativa de alcance general y eficacia directa como los Reglamentos, o el desarrollo de algunos principios clave como la supervisión humana o la discriminación.

ABSTRACT

On 20 October 2020, the European Parliament adopted a resolution (2020/2012(INL)) with recommendations to the Commission regarding artificial intelligence, robotics and related technologies, which included a legislative proposal for a Regulation on the ethical principles for the development, deployment and use of these technologies. The content of this proposal undoubtedly follows from the regulatory vision that the European Commission has maintained in documents such as the White Paper on Artificial Intelligence (COM(2020) 65 final) or the Ethical guidelines for trustworthy AI drawn up by the High-Level Expert Group on AI. Given this new legislative horizon, it is more necessary than ever to address a constructive criticism on the proposal, highlighting the possibility of reformulating its markedly soft-law character despite its location in a regulatory source of general application and directly applicable, such as regulations, or the adopted approach for certain key principles such as human supervision or discrimination.

KEYWORDS

Artificial intelligence
European Union
Algorithms
Discrimination
Human supervision

I. INTRODUCCIÓN

El pasado 20 de octubre de 2020, el Parlamento Europeo (Parlamento o PE, en adelante) aprobó la Resolución (2020/2012(INL)) con recomendaciones destinadas a la Comisión sobre inteligencia artificial, robótica y tecnologías conexas¹, en cuyo anexo primero se recoge una propuesta legislativa para la tramitación de un Reglamento sobre los principios éticos para el desarrollo, el despliegue y el uso de dichas tecnologías. La iniciativa, preparada por el parlamentario Ibán García del Blanco (S&D, España), fue aprobada por 559 votos a favor, 44 en contra y 88 abstenciones. Con esta resolución, la Unión Europea continúa consolidando el camino hacia una regulación específica para las tecnologías que incorporan inteligencia artificial, un camino transitado en los últimos años por documentos elaborados en el seno de la Comisión Europea (Comisión o CE), así como por el trabajo desarrollado por el Servicio de Estudios Parlamentarios (EPRS) y el Grupo de Expertos sobre el Futuro de la Ciencia y la Tecnología (STOA), cuyo relevo toma ahora el Parlamento.

1. Resolución del Parlamento Europeo, de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas (2020/2012(INL)). Disponible en: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_ES.html

En lo que respecta al Reino de España, la aprobación de esta resolución coincide temporalmente con la reciente consulta pública abierta por la Secretaría de Estado de Digitalización e Inteligencia Artificial del Ministerio de Asuntos Económicos y Transformación Digital sobre la Carta de Derechos Digitales de la ciudadanía², así como con la publicación de la Estrategia Nacional de Inteligencia Artificial (ENIA) elaborada por el Grupo de Trabajo Interministerial en Inteligencia Artificial, coordinado por el Ministerio de Ciencia e Innovación³, y para la que se ha anunciado inversión pública de 600 millones en el periodo 2021-2023⁴.

En este texto se analizarán los contenidos básicos de esta propuesta que, sin lugar a duda, deriva de la visión regulatoria que la Comisión había mantenido en documentos como el Libro Blanco sobre Inteligencia Artificial⁵ o en las Directrices éticas para una IA fiable elaboradas por el Grupo independiente de expertos de alto nivel⁶. Para realizar este análisis se han seleccionado algunos de los aspectos más relevantes recogidos por la propuesta – ahora concebidos como principios⁷ – y que, a su vez, ya venían desarrollados por estos dos documentos: el enfoque basado en el riesgo para la regulación de la IA, la supervisión e intervención humana como mecanismo de gobernanza y la protección frente a los riesgos discriminatorios ante los sesgos que introducen estas tecnologías.

2. La consulta se cerró el pasado 4 de diciembre de 2020. El documento para consulta pública está disponible en el siguiente enlace: https://portal.mineco.gob.es/RecursosArticulo/mineco/ministerio/participacion_publica/audiencia/ficheros/SEDIACartaDerechosDigitales.pdf [fecha de la última consulta: 5 de diciembre de 2020]

3. Estrategia Nacional de Inteligencia Artificial (ENIA). Gobierno de España. Versión 1.0. Disponible en: <https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/021220-ENIA.pdf> [fecha de la última consulta: 3 de diciembre de 2020]

4. “Pedro Sánchez presenta la Estrategia Nacional de Inteligencia Artificial con una inversión pública de 600 millones en el periodo 2021-2023”. Nota de prensa del Gobierno de España. Disponible en: <https://www.lamoncloa.gob.es/presidente/actividades/Paginas/2020/021220-sanchezenia.aspx> [fecha de la última consulta: 3 de diciembre de 2020]

5. Vid. COMISIÓN EUROPEA (CE), “Libro Blanco sobre la inteligencia artificial - un enfoque europeo orientado a la excelencia y la confianza”, Bruselas, COM(2020) 65 final, 19 de febrero de 2020. Disponible en: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_es.pdf

6. Vid. GRUPO DE EXPERTOS DE ALTO NIVEL SOBRE IA (HLEG-AI), “Directrices éticas para una IA fiable”, Bruselas, 8 de abril de 2019. Disponible en: <https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1>

7. Anexo a la Resolución. A. Principios y objetivos de la propuesta solicitada. III. El «Reglamento sobre principios éticos para el desarrollo, el despliegue y el uso de la inteligencia artificial, la robótica y las tecnologías conexas» se basa en los siguientes principios: – una inteligencia artificial, una robótica y unas tecnologías conexas antropocéntricas, antropogénicas y controladas por seres humanos; – evaluación de la conformidad obligatoria de la inteligencia artificial, la robótica y las tecnologías conexas de alto riesgo; – seguridad, transparencia y rendición de cuentas; – salvaguardias y vías de recurso contra el sesgo y la discriminación; – derecho de resarcimiento; – responsabilidad social e igualdad de género en la inteligencia artificial, la robótica y las tecnologías conexas; – una inteligencia artificial, una robótica y unas tecnologías conexas ambientalmente sostenibles; – respeto de la intimidad y restricciones al uso del reconocimiento biométrico; – buena gobernanza relativa a la inteligencia artificial, la robótica y las tecnologías conexas, incluidos los datos utilizados o producidos por dichas tecnologías.

Por último, se incluye un apartado en el que se valora el carácter y elementos de la propuesta como norma jurídica.

II. APLICACIONES DE ALTO RIESGO

El enfoque basado en el riesgo es una característica fundamental sobre la que se cimenta la estructura de esta propuesta. La Comisión recogía ya este enfoque en el Libro Blanco sobre IA⁸ declarando la necesidad de que el marco regulador de estas tecnologías fuese proporcional, en el sentido de que el nivel de intervención regulatoria debe ajustarse al nivel de riesgo que representan las distintas aplicaciones y usos de inteligencia artificial. Para el cumplimiento de este objetivo, la Comisión declara la necesidad de establecer criterios claros en la normativa a la hora de determinar el riesgo que entrañan las distintas tecnologías⁹. Siguiendo esta misma lógica, en la introducción de la resolución el Parlamento declara la necesidad de adoptar este enfoque para no obstaculizar las innovaciones en el futuro ni crear cargas innecesarias; y justifica, asimismo, su acción normativa sobre la necesidad de garantizar una aplicación homogénea de principios éticos comunes para las aplicaciones de calificadas alto riesgo.

Este enfoque normativo basado en el riesgo está ya presente en otras regulaciones europeas; entre otros, en el Reglamento General de Protección de Datos (RGPD¹⁰) – un ejemplo claro es la obligatoriedad contenida en su artículo 35 de llevar a cabo una evaluación de impacto de protección de datos en función del riesgo del tratamiento¹¹ – o en el Reglamento de Productos Sanitarios (MDR¹²) – en las reglas de clasificación de los productos sanitarios contenidos en el artículo 51 y anexos de dicha norma¹³ –.

El Libro Blanco sobre IA estableció un sistema cumulativo por el que establecer si determinada aplicación podía considerarse o no de alto riesgo y, por ende, aplicarse los

8. La referencia a este enfoque fue residual en las Directrices del HLEG-AI. En este documento las referencias a un enfoque basado en el riesgo se refieren a las preocupaciones que pudieren surgir a largo plazo y, en particular, la necesidad de tener en cuenta preocupaciones hipotéticas, en vista de los posibles «desconocidos desconocidos» y «cisnes negros» - que define como sucesos muy poco frecuentes, aunque con elevado impacto; *tan raro, que puede que ni siquiera sea observado* -. GRUPO DE EXPERTOS DE ALTO NIVEL SOBRE IA (HLEG-AI), "Directrices éticas para una IA fiable", op cit., p. 46

9. COMISIÓN EUROPEA (CE), "Libro Blanco sobre la inteligencia artificial - un enfoque europeo orientado a la excelencia y la confianza", op cit., p. 21.

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

11. Vid. DEMETZOU, K., "Data Protection Impact Assessment: A Tool for Accountability and the Unclear Concept of 'High Risk' in the General Data Protection Regulation.", *Computer Law & Security Review*, vol. 35, núm. 6, 2019. <https://doi.org/10.1016/j.clsr.2019.105342>

12. Reglamento (UE) 2017/745 del Parlamento Europeo y del Consejo, de 5 de abril de 2017, sobre los productos sanitarios, por el que se modifican la Directiva 2001/83/CE, el Reglamento (CE) n.º 178/2002 y el Reglamento (CE) n.º 1223/2009 y por el que se derogan las Directivas 90/385/CEE y 93/42/CEE del Consejo

13. Vid. KISELEVA, A., "AI as a Medical Device: Is It Enough to Ensure Performance Transparency and Accountability?", *European Pharmaceutical Law Review*, vol. 4, núm. 1, 2020

requisitos obligatorios establecidos por el documento¹⁴. Los criterios para evaluar dicho riesgo son¹⁵, en primer lugar, que por la actividad habitual del sector al que pertenezca la aplicación sea previsible que existan riesgos significativos – estableciendo como ejemplos la sanidad, el transporte, la energía y determinados ámbitos del sector público – y, en segundo lugar, que del uso concreto de la aplicación en dicho sector puedan surgir igualmente riesgos significativos, siendo necesario evaluar las posibles repercusiones de este uso – por ejemplo, la producción de efectos jurídicos o similares, el riesgo de causar lesiones, la muerte, o daños materiales o inmateriales significativos o efectos que no puedan evitarse razonablemente¹⁶ –. No parece que los criterios resulten lo suficientemente claros para aportar la seguridad jurídica necesaria en la implementación de este enfoque¹⁷.

En cuanto a la Resolución del Parlamento, el enfoque basado en el riesgo se desarrolla con mayor detalle, permitiendo una mejor comprensión del concepto de “alto riesgo” que es, a fin de cuentas, el concepto que adquiere una entidad jurídica de capital relevancia dentro de este enfoque. En su Considerando 11, la propuesta establece que una IA de alto riesgo es aquella que contiene *un riesgo significativo de causar lesiones o daños a particulares o a la sociedad, vulnerando los derechos fundamentales y las normas de seguridad establecidas en el Derecho de la Unión*. Para evaluar este riesgo hace referencia a la necesidad de tener en cuenta tres factores, también así recogidos en la evaluación de riesgos del artículo 14: (1) el sector al que pertenece, (2) el uso o finalidad específica al que se dedica y (3) la gravedad del daño que cabe esperar que produzca. En comparación con el Libro Blanco sobre IA, podemos ver que se mantiene el primer criterio, mientras que el segundo es dividido en otros dos criterios diferenciados, uso o finalidad y gravedad del daño.

Esta evaluación es determinante dado que, en analogía al Libro Blanco sobre IA con los requisitos obligatorios, la consideración de una tecnología de alto riesgo implica que estén sujetas las obligaciones jurídicas y los principios éticos del marco regulador (artículo 6 de la propuesta de Reglamento). Por ello, la resolución indica la necesidad de realizar una evaluación *ex ante imparcial, regulada y externa apoyada en criterios*

14. A saber, datos de entrenamiento; datos y registros de datos; información que debe facilitarse; solidez y exactitud; supervisión humana; requisitos específicos en el caso de determinadas aplicaciones de IA, como las empleadas para la identificación biométrica remota.

15. COMISIÓN EUROPEA (CE), “Libro Blanco sobre la inteligencia artificial - un enfoque europeo orientado a la excelencia y la confianza”, op cit., pp. 22 y 23

16. Este parece el único vínculo entre el concepto de enfoque basado en el riesgo manejado por las Directrices del HLEG-AI y el Libro Blanco de la Comisión.

17. A juzgar por el proceso de consulta pública de este documento, dicho concepto parece no estar claro para la ciudadanía, ya que sólo un 37% de los encuestados en la consulta pública abierta sobre el Libro Blanco respondieron a la pregunta: “¿Está usted de acuerdo con el enfoque para determinar las aplicaciones de IA de “alto riesgo” propuestas (...)?” (traducción por el autor). Vid. Informe resumido de la consulta pública abierta sobre el Libro Blanco sobre Inteligencia Artificial. Disponible en: <https://ec.europa.eu/digital-single-market/en/news/white-paper-artificial-intelligence-public-consultation-towards-european-approach-excellence> [fecha de la última consulta: 28 de noviembre de 2020]

concretos y definidos y, para ello, elabora una lista exhaustiva y acumulativa de sectores de alto riesgo y de usos o fines de alto riesgo (Vid. Tabla 1)¹⁸. En virtud del apartado tercero del artículo 14, la lista se elaborará y actualizará por la Comisión en colaboración con las autoridades nacionales de control creadas por el artículo 18 de esta misma propuesta.

Si la lista exhaustiva y acumulativa parece normativizar con nítida seguridad jurídica los criterios para determinar cuándo una tecnología pertenece a un sector de alto riesgo y contiene un uso o finalidad de alto riesgo, no parecen tan nítidos los criterios para determinar la gravedad del daño que cabe esperar que produzca dicha tecnología. Las únicas referencias al respecto las encontramos en la definición de lesión o daño del artículo cuarto¹⁹, así como en el Considerando 11 – que no tendría valor jurídico, si interpretativo – cuando indica que dicha gravedad se determina en función de la *magnitud de la lesión o daño potencial, el número de personas afectadas, el valor total del perjuicio ocasionado y el daño a la sociedad en su conjunto*. Desde luego resultaría oportuno que, al menos, estos subcriterios para determinar la gravedad del daño figuraran también en el articulado del Reglamento.

Conforme al apartado segundo del artículo 14, la evaluación de riesgos que deberá determinar si una tecnología es de alto riesgo, se realizará por las autoridades nacionales de control arriba mencionadas, bajo la coordinación de la Comisión Europea, así como de cualquier institución de la Unión que pueda designarse a tal fin. Estas mismas autoridades independientes de control serán responsables a su vez de realizar la evaluación de conformidad con las obligaciones establecidas en el Reglamento para las tecnologías de alto riesgo conforme a lo dispuesto en el artículo 15. En definitiva, en el análisis del enfoque normativo basado en el riesgo, observamos que la propuesta del Parlamento desarrolla los criterios adelantados por la Comisión en el Libro Blanco sobre IA, y es destacable el papel determinante que otorga a estas autoridades nacionales de control a la hora de garantizar una evaluación de riesgos *ex ante imparcial, regulada y externa*.

18. Aunque la lista recoge, por lo general, sectores y usos que ya aparecen en el Libro Blanco sobre IA, resulta llamativa la inclusión expresa del sector de seguridad y defensa y del uso de sistemas militares autónomos en dicha lista, dado que los fines militares fueron expresamente excluidos en el libro. COMISIÓN EUROPEA (CE), “Libro Blanco sobre la inteligencia artificial - un enfoque europeo orientado a la excelencia y la confianza”, op cit., p.1

19. artículo 4 n) «lesión o daño», toda lesión física o mental y todo daño material o inmaterial, incluso cuando hayan sido causados por incitación al odio, sesgo, discriminación o estigmatización, tales como pérdidas financieras o económicas, pérdida de empleo o de oportunidades educativas, restricción indebida de la libertad de elección o de expresión, pérdida de la privacidad y cualquier infracción del Derecho de la Unión que sea perjudicial para una persona

Tabla 1

<p>Sectores de alto riesgo</p>	<ul style="list-style-type: none"> — Empleo — Educación — Asistencia sanitaria — Transporte — Energía — Sector público (asilo, migración, controles fronterizos, sistema judicial y servicios de seguridad social) — Seguridad y defensa — Finanzas, bancos, seguros
<p>Usos o fines de alto riesgo</p>	<ul style="list-style-type: none"> — Contratación — Clasificación y evaluación de estudiantes — Asignación de fondos públicos — Concesión de préstamos — Comercio, corretaje, fiscalidad, etc. — Tratamientos y procedimientos médicos — Procesos electorales y campañas políticas — Decisiones del sector público que tienen un impacto significativo y directo en los derechos y las obligaciones de las personas físicas o jurídicas — Conducción automatizada — Gestión del tráfico — Sistemas militares autónomos — Producción y distribución de energía — Gestión de residuos — Control de emisiones

III. SUPERVISIÓN HUMANA INTEGRAL

La reivindicación de una IA centrada en el ser humano ha ocupado siempre un papel determinante en la visión regulatoria de las instituciones europeas, y por ello esta propuesta del Parlamento exige el desarrollo de una IA antropocéntrica y antropogénica²⁰ en su artículo séptimo, cuyo contenido puede resumirse en la garantía de una supervisión humana integral (apartado primero) que, en todo caso, permita el restablecimiento del control humano cuando sea necesario (apartado segundo). Si tenemos en cuenta que la supervisión humana aparece como uno de los siete requerimientos del HLEG-AI para el desarrollo de una IA fiable y como uno de los requisitos de obligado cumplimiento para

20. Aunque en el texto de la resolución sí desarrolla lo antropocéntrico y antropogénico de la IA (1. a 11. del texto de la resolución), en realidad, muchos de estos aspectos no están necesariamente relacionados con la supervisión humana integral, sin embargo, el artículo 7 que recoge como principio *una inteligencia artificial, una robótica y unas tecnologías conexas antropocéntricas, antropogénicas y controladas por seres humano*, tan solo recoge la regulación de la supervisión y control humano como mecanismo de gobernanza, es por ello que este artículo centra el análisis de este artículo en dicho mecanismo de gobernanza.

las aplicaciones de alto riesgo en el Libro Blanco sobre IA, no sorprende la inclusión de ésta en la propuesta aquí analizada.

A la hora de determinar las distintas formas de supervisión e intervención humana en el contexto automatizado en el que operan estas tecnologías, es necesario considerar cómo se concibe este contexto en sí. En este sentido, COBBE y SINGH abogan por conceptualizar estas tecnologías desde un marco más amplio que el mero estadio en el que se genera una decisión, o más amplio que un mero sistema de toma de decisiones, más como un sistema que se compone tanto de aspectos sociales como técnicos y que opera en un contexto institucional concreto²¹. La doctrina, desde muy diferentes ámbitos y perspectivas, ha tratado de atomizar los procesos técnico-sociales de toma de decisiones en relación con la IA, definiendo distintas fases o etapas de relevancia para el análisis normativo y ético de la cuestión. Siguiendo este mismo esquema, la propuesta del Parlamento considera relevantes tres fases que forman parte del título de la propuesta y define en su artículo cuarto: desarrollo, despliegue y uso²². En la siguiente tabla puede observarse el esquema seguido por el Parlamento en comparación con otras aportaciones relevantes de la doctrina.

Los documentos de la Comisión y el HLEG-AI coinciden en que el tipo y nivel adecuado de supervisión humana puede variar, traduciéndose en la posibilidad de establecer diferentes mecanismos de gobernanza. La participación o intervención humana se produce tanto en las fases de desarrollo como en las posteriores, y la entidad de dicha intervención no tiene por qué ser más garantista por producirse en la fase de implementación o despliegue²³. La propuesta de Reglamento considera oportuno enlazar las fases definidas con un rol humano determinado (vid. Tabla 2), definiendo igualmente estos roles en el artículo cuarto: desarrollador, desplegador y usuario²⁴. Es habitual la

21. COBBE, J. y SINGH, J., "Reviewable Automated Decision-Making," *Computer Law & Security Review*, vol. 39, 2020, p. 2.

22. Artículo 4: (...) f) «desarrollo», la construcción y el diseño de algoritmos, la escritura y el diseño de programas informáticos o la recopilación, el almacenamiento y la gestión de datos con el fin de crear o entrenar la inteligencia artificial, la robótica y las tecnologías conexas o de crear una nueva aplicación para la inteligencia artificial, la robótica y las tecnologías conexas existentes; (...) h) «despliegue», el funcionamiento y la gestión de la inteligencia artificial, la robótica y las tecnologías conexas, así como su comercialización o cualquier otra forma de puesta a disposición de los usuarios; (...) j) «uso»: toda acción relacionada con la inteligencia artificial, la robótica y las tecnologías conexas distinta del desarrollo o el despliegue;

23. ALMADA, M., "Human Intervention in Automated Decision-Making: Toward the Construction of Contestable Systems," *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*, 2019, p. 3.

24. Artículo 4: (...) g) «desarrollador», toda persona física o jurídica que toma decisiones que determinan y controlan el curso o la forma del desarrollo de la inteligencia artificial, la robótica y las tecnologías conexas; (...) i) «desplegador», toda persona física o jurídica que interviene en el despliegue específico de la inteligencia artificial, la robótica y las tecnologías conexas con una función de control o gestión, tomando decisiones, ejerciendo un control sobre los riesgos y beneficiándose de dicho despliegue; (...) k) «usuario»: toda persona física o jurídica que utiliza la inteligencia artificial, la robótica y las tecnologías conexas para fines distintos del desarrollo o el despliegue;

relación entre los distintos grados de autonomía y la intervención humana, de hecho, el Considerando 11 menciona la norma SAE J3016 para la conducción automatizada que propone seis niveles de automatización²⁵, aunque es importante reseñar que estas clasificaciones o niveles no son el resultado de una reducción “cuantitativa” per sé de la intervención humana, sino más bien una forma de calificar la forma en que la agencia y el control humanos operan en un contexto automatizado.

Tabla 2

De Laat	Cobbe y Singh		Propuesta de Reglamento
Fases	Etapas	Fases	Fases
(...)	Adquisición	Encargo	(...)
	Definición del problema		
Recolección de Datos	Recolección de Datos	Construcción del modelo	Desarrollo del modelo (desarrollador)
Construcción del modelo	Pre-procesado		
	Entrenamiento del modelo		
	Testeo del modelo		
Uso del modelo	Despliegue	Toma de decisiones	Despliegue del modelo (desplegador)
	Uso		Uso del modelo (usuario)
	Consecuencias		
(...)	Auditoría	Investigación	(...)
	Revelación		

Acerca del contenido concreto del artículo 7, en su apartado primero establece que debe de garantizarse en todo momento, haciendo referencia a las fases de desarrollo, despliegue y uso, una supervisión humana integral. No opta, en definitiva, por ningún mecanismo de intervención humana en particular, sino que se limita a regular que la supervisión humana debe darse en todo el ciclo de vida de la IA. El Considerando 10 sí incluye una referencia algo más concreta, declarando que las decisiones adoptadas deben ser objeto de revisión, evaluación, intervención y control humanos significativos; es decir, menciona cuatro formas de participación humana (revisión, evaluación, intervención y control) y hace referencia al concepto “significativo” (*meaningful* en la literatura anglófona), que aparece también en la resolución del Parlamento en relación

25. SAE J3016, actualizada por última vez en 2018 como J3016_201806. Disponible en: https://www.sae.org/standards/content/j3016_201806/ [fecha de la última consulta: 9 de diciembre de 2020]

a los sistemas de armas autónomos letales (SAAL) y a la protección de la intimidad en la toma de decisiones por los poderes públicos.

Aunque cuenta con un desarrollo mayor en el ámbito militar dado el debate surgido en el derecho internacional por el despliegue de los SAAL, sobre este concepto no existe actualmente un desarrollo teórico suficiente sobre el que las personas desarrolladoras puedan adoptar decisiones con seguridad jurídica²⁶, si bien, ello no ha impedido que el Grupo de Trabajo sobre Protección de Datos del Artículo 29 haga referencia a esa significancia en la toma de decisiones automatizada regulada en el artículo 22 del RGPD²⁷. Podría resultar relevante añadir a este artículo 7 de la propuesta dicho concepto, de forma que deba garantizarse una supervisión humana integral y significativa, de forma que haya referencia no solo a las fases en que debe producirse la supervisión (integral), sino también a la cualidad de ésta (significativa)²⁸.

Analizando el contenido del apartado segundo, puede observarse que, aquí sí, el Parlamento Europeo considera necesario incluir un mecanismo de gobernanza específico que permita el restablecimiento del control humano cuando sea necesario, *incluso mediante la alteración o la desactivación de dichas tecnologías*. Esta clase de mecanismo tiene más sentido para aquellas tecnologías cuyo despliegue se produce en una dimensión física y, tal y como se desarrolla en el Libro Blanco sobre IA, cabe el restablecimiento del control tanto desde su diseño en el desarrollo del sistema, como desde la supervisión o vigilancia en tiempo real por parte de operadores humanos en las fases de despliegue y uso²⁹. Los ejemplos utilizados por el Libro Blanco sobre IA son clarificadores a este respecto³⁰:

- *Se realiza un seguimiento del sistema de IA mientras funciona y es posible intervenir en tiempo real y desactivarlo (por ejemplo, un vehículo sin conductor cuenta con un procedimiento o botón de apagado para las situaciones en las que un humano determine que el funcionamiento del vehículo no es seguro).*
- *En la fase de diseño, se imponen restricciones operativas al sistema de IA (por ejemplo, un vehículo sin conductor dejará de funcionar en determinadas condiciones de visibilidad reducida en las que los sensores sean menos fiables, o mantendrá una cierta distancia con el vehículo que lo preceda en una situación dada).*

26. Vid. De SIO, F. y Van Den HOVEN, J., "Meaningful Human Control over Autonomous Systems: A Philosophical Account.", *Frontiers in Robotics and AI*, vol. 5, 2018, pp. 1-15

27. Vid. GRUPO DE TRABAJO SOBRE PROTECCIÓN DE DATOS DEL ARTÍCULO 29 (GT29), "Directrices Sobre Decisiones Individuales Automatizadas y Elaboración de Perfiles a Los Efectos Del Reglamento 2016/679.", Bruselas, 2018, pp. 1-37. Disponible en: <https://www.aepd.es/sites/default/files/2019-12/wp251rev01-es.pdf>

28. Aunque no puede obviarse que es necesario un desarrollo teórico de este concepto para que resulte operativo.

29. FISCHER, J.E., GREENHALGH, C., JIANG, W., RAMCHURN, S.D., WU, F. y RODDEN, T., "In-the-loop or on-the-loop? Interactional arrangements to support team coordination with a planning agent.", *Concurrency and Computation: Practice and Experience*, 2017, p. 1.

30. COMISIÓN EUROPEA (CE), "Libro Blanco sobre la inteligencia artificial - un enfoque europeo orientado a la excelencia y la confianza", op cit., p. 26.

IV. SESGOS Y DISCRIMINACIÓN

No hay duda de que la discriminación es una de las principales preocupaciones en torno al uso de la inteligencia artificial. La problemática deriva de que la propia función de los algoritmos de aprendizaje automático³¹ es discriminar entre los datos con los que se genera su aprendizaje, discriminar en el sentido de seleccionar excluyendo, y con dicho objetivo se utilizan precisamente, lo cual no quiere decir que las formas en las que discriminan sean siempre aceptables por nuestro sistema jurídico, aunque puedan ser válidas en un sentido estadístico o predictivo³². La Comisión Europea ha considerado que deben adoptarse medidas razonables para velar porque el uso posterior de estos sistemas no genere discriminación, para lo cual considera la necesidad de adoptar medidas razonables como utilizar conjuntos de datos que sean suficientemente representativos o la necesidad de conservar documentación sobre metodologías de programación y entrenamiento³³.

La preocupación del Parlamento por esta cuestión parece ir mucho más allá, y no solo porque la ausencia de sesgos y discriminación haya sido configurada como un principio ético de obligado cumplimiento para las tecnologías de alto riesgo. La resolución muestra una clara vinculación entre las *data-driven technologies* (hace referencia a los conjuntos de datos en investigación y desarrollo, así como a los datos de entrenamiento) y la creación y reproducción de sesgos que, a su vez, aumenta el riesgo de discriminación. Asimismo, la discriminación está concebida como una de las posibles causas en la producción de “lesión o daño” por estas tecnologías en el artículo 4 de la propuesta, y el derecho a la no discriminación está presente en toda la propuesta como uno de los pilares básicos del Derecho de la Unión.

Acerca de las definiciones que adopta sobre sesgo y discriminación³⁴, es reseñable que la definición de discriminación es puramente normativa sobre la base del Derecho de la Unión, mientras que el término sesgo se define como un prejuicio desde un plano ético y no tanto estadístico, es decir, no se hace referencia a la concepción de sesgo como una desviación estadística. La introducción de un sesgo estadístico en este sentido puede incluso representar una compensación de un resultado que se considera sesgado en el sentido de éticamente problemático³⁵. Es posible que la definición de sesgo como un prejuicio o percepción personal sea un tanto limitada en este contexto y, por ello, considero que podría adoptarse una definición más ajustada a partir la consideración

31. El campo que mayor protagonismo ha acaparado en el actual interés por la IA y que constituye el paradigma de las *data-driven technologies* o tecnologías basadas en datos.

32. VEALE, M. y BINNS, R., “Fairer Machine Learning in the Real World: Mitigating Discrimination without Collecting Sensitive Data.”, *Big Data & Society*, vol. 4, núm. 2, 2017, p. 2

33. COMISIÓN EUROPEA (CE), “Libro Blanco sobre la inteligencia artificial - un enfoque europeo orientado a la excelencia y la confianza”, op cit., pp. 23 y 24.

34. Artículo 4: l) «sesgo», toda percepción personal o social prejuiciosa de una persona o de un grupo de personas sobre la base de sus características personales; m) «discriminación», todo trato diferenciado de una persona o de un grupo de personas basado en un motivo que no tiene justificación objetiva o razonable alguna y que, por tanto, está prohibido por el Derecho de la Unión.

35. Vid. DANKS, D. y LONDON, A.J., “Algorithmic Bias in Autonomous Systems.” *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, AAAI Press, 2017, pp. 4691–4697.

que realiza HILDEBRANDT³⁶ sobre los sesgos en el plano ético, la cual podría ser: l) «sesgo», toda desviación o representación estadística que reconfigura la distribución de los bienes, servicios, riesgos y oportunidades de una o varias personas físicas o jurídicas.

El contenido del apartado primero es significativamente exhaustivo en cuanto a las categorías discriminatorias que enumera, a saber: *no discriminarán por motivos de raza, sexo, orientación sexual, embarazo, discapacidad, características físicas o genéticas, edad, minoría nacional, origen étnico o social, lengua, religión o creencias, opiniones políticas o participación cívica, nacionalidad, estado civil o económico, educación o antecedentes penales*. En el apartado segundo se hace referencia a los criterios para la justificación de un trato desigual, de acuerdo con la jurisprudencia y el Derecho de la Unión.

La lista tan exhaustiva que elabora la propuesta podría estar relacionada con una de las características inherentes al uso de algoritmos de aprendizaje automático, dado que el análisis inferencial de los mismos, esto es, la capacidad de establecer inferencias a partir de patrones y correlaciones ocultas, amplía la gama de víctimas de actos discriminatorios, víctimas que podrían no corresponderse con las categorías previstas por la normativa y, por ende, el Derecho antidiscriminatorio no protegería adecuadamente contra estos riesgos³⁷. Es decir, la lista exhaustiva del Parlamento podría representar un intento de cubrir el máximo número posible de categorías vulnerables al uso de estas tecnologías. Ahora bien, estos análisis inferenciales se caracterizan, precisamente, por la creación de grupos *ad hoc*³⁸ que podrían sufrir igualmente efectos perjudiciales y que no podrían enumerarse en una lista por muy exhaustiva que fuere.

El notable consenso en la doctrina acerca de lo injustas que pueden resultar determinadas inferencias algorítmicas que evaden la normativa antidiscriminatoria, hasta el momento, en la mayoría de las ocasiones no se ha traducido, sin embargo, en recomendaciones político-jurídicas concretas para paliar estos efectos. La reciente propuesta ZUIDERVEEN aboga por nuevas normas que aborden desde una perspectiva sectorial los efectos discriminatorios particulares – y no debidamente abarcados por la normativa antidiscriminatoria – que se produzcan en cada uno de esos sectores³⁹, con lo que a la propuesta del Parlamento que apuesta por una regulación común de los principios éticos

36. HILDEBRANT, M., "The issue of bias: the framing powers of ML", Draft version, p. 1, en Marcello Pelillo, Teresa Scantamburlo (eds.), *Machine Learning and Society: Impact, Trust, Transparency*, MIT Press forthcoming 2020

37. WACHTER, S. "Affinity Profiling and Discrimination by Association in Online Behavioural Advertising." *Berkeley Technology Law Journal*, vol. 35, núm. 2, 2020, Forthcoming, p. 56.

38. V. más ampliamente sobre este concepto MITTELSTADT, B., "From Individual to Group Privacy in Big Data Analytics.", *Philosophy & Technology*, vol. 30, núm. 4, 2017, pp. 475–94, doi:10.1007/s13347-017-0253-7. También se ha definido la elaboración de perfiles como un proceso de construcción de conocimientos a partir de grandes conjuntos de datos que, a menudo, no guardan relación directa con los fundamentos protegidos por el derecho antidiscriminatorio. Vid. MANN, M. y MATZNER, T., "Challenging Algorithmic Profiling: The Limits of Data Protection and Anti-Discrimination in Responding to Emergent Discrimination.", *Big Data & Society*, vol. 6, núm. 2, 2019, pp. 1–11.

39. ZUIDERVEEN BORGESIU, F.J., "Strengthening Legal Protection against Discrimination by Algorithms and Artificial Intelligence.", *The International Journal of Human Rights*, vol. 24, núm. 10, 2020, p. 15.

de la IA a todos los sectores, difícilmente podríamos exigirle que integrase esta perspectiva como solución a los riesgos discriminatorios no abarcados por la normativa actual.

V. REGULACIÓN DE LOS PRINCIPIOS ÉTICOS

Ha de reconocerse que el título escogido para la propuesta es, cuando menos, llamativo: *Reglamento sobre principios éticos para el desarrollo, el despliegue y el uso de la inteligencia artificial, la robótica y las tecnologías conexas*. Esta cristalización de unos principios éticos, así denominados expresamente, en una fuente normativa de alcance general y eficacia directa como los Reglamentos, no es habitual. Mis recelos, prejuicios o sesgos, si se prefiere, ante esta denominación derivan de algunas de las críticas que MITTELSTADT formuló a partir del análisis de nada menos que 84 declaraciones institucionales, públicas y privadas, que describen principios, valores y otros fundamentos de alto nivel para guiar el desarrollo ético, el despliegue y la gobernanza de la IA⁴⁰. También derivan, en parte, por la estrategia de *ethics washing* que empresas multinacionales tecnológicas han mantenido en los últimos años⁴¹ y que está de actualidad tras el reciente despido de Timnit Gebru del equipo de Inteligencia Artificial Ética de Google que codirigía⁴².

La forma adecuada para neutralizar estos sesgos con los que abordo este análisis, no puede ser otra que observar con detenimiento los mecanismos de gobernanza que esta propuesta establece para hacer efectiva la aplicación de esos principios éticos que transforma en obligaciones jurídicas para las tecnologías de IA de alto riesgo.

Se han mencionado ya previamente las autoridades de control que esta propuesta pretende crear con las siguientes funciones principales: (1) evaluación del riesgo de las tecnologías en el ámbito de aplicación de la propuesta para la calificación, en su caso, como tecnologías de alto riesgo de conformidad con el artículo 14; (2) evaluación – y supervisión – de la conformidad de las tecnologías calificadas como de alto riesgo con los principios establecidos en los artículos 6 a 12, incluyendo la supervisión humana integral y la ausencia de sesgos y discriminación; (3) emisión del certificado europeo de conformidad ética para las IA de alto riesgo evaluadas, y al que también pueden acceder voluntariamente desarrolladores de IA que no sean calificadas de alto riesgo; (4) ser responsable de la aplicación de las normas para la gobernanza establecidas por la propuesta y, en particular, servir de primer punto de contacto en caso de sospecha de incumplimiento de las obligaciones jurídicas y los principios éticos. Estas funciones deben realizarse en cooperación con las autoridades de control de los demás Estados

40. Me estoy refiriendo a MITTELSTADT, B. 2019. "Principles Alone Cannot Guarantee Ethical AI.", *Nature Machine Intelligence*, vol. 1, núm. 11, 2019, pp. 501–507.

41. "In 2020, let's stop AI ethics-washing and actually do something", Karen Hao en *MIT Technology Review*, 27 de diciembre de 2019: <https://www.technologyreview.com/2019/12/27/57/ai-ethics-washing-time-to-act/> [fecha de la última consulta: 10 de diciembre de 2020]

42. Aquí la plataforma en su apoyo que se ha creado: "Standing with Dr. Timnit Gebru", *Medium*, 4 de diciembre de 2020: <https://googlewalkout.medium.com/standing-with-dr-timnit-gebru-isupporttimnit-believeblackwomen-6dad300d382> [fecha de la última consulta: 10 de diciembre de 2020]

miembros, así como con otras autoridades responsables de la aplicación de la legislación sectorial, la Comisión o cualquier institución pertinente de la Unión.

El establecimiento de una autoridad pública de supervisión y control parece un avance notable en relación al estado actual del ordenamiento⁴³, especialmente en ámbitos que quedan al margen de normas sectoriales que establecen otras autoridades de control. Ahora bien, no puede dejarse al margen la limitación fundamental que se ha establecido, por omisión, en las funciones de estas autoridades de control: no se establece un régimen sancionador para el incumplimiento por tecnologías calificadas de alto riesgo de las obligaciones establecidas en los artículos 6 a 12. Cualquier otra consideración respecto de la forma en la que se establecen las evaluaciones y certificaciones son, ciertamente, secundarias ante esta omisión.

Como consideración final, sobre la figura de las autoridades de control conviene advertir del riesgo que pueden correr si no reciben una financiación y potestades jurídicas necesarias para el cumplimiento de sus funciones, incluyendo la capacidad para asumir la carga de trabajo que la actividad regulada requiere. Y es que, el establecimiento de un régimen sancionador es condición necesaria pero no suficiente para asegurar el cumplimiento normativo, tal y como se ha puesto de manifiesto en el marco de la protección de datos personales en múltiples ocasiones. Así lo expresa ZUIDERVEEN en relación al control de la no discriminación: *"Equality Bodies and Data Protection Authorities should have sufficient investigation and enforcement powers, and should receive adequate funding, for instance to hire technical expertise. In conclusion, new regulation should aim for better enforcement of current non-discrimination norms in the area of algorithmic decision-making"*⁴⁴.

VI. CONCLUSIONES

En este texto ha sido analizada la propuesta de Reglamento sobre los principios éticos para el desarrollo, el despliegue y el uso de la inteligencia artificial, la robótica y las tecnologías conexas, aprobada por el Parlamento Europeo en la Resolución 2020/2012(INL) del pasado 20 de octubre. El análisis se ha centrado sobre algunos aspectos destacados de la propuesta que han sido seleccionados, o bien por el interés que la doctrina ha mantenido sobre la regulación de los mismos, o bien por la importancia que estos aspectos han adquirido en base a la propia propuesta.

Se han trabajado y desarrollado conceptos como "alto riesgo", "supervisión humana", "sesgos" o "discriminación" que representan, sin lugar a duda, complicados retos regulatorios que el Parlamento se ha propuesto afrontar. Este reto asumido por el Parlamento

43. Acerca de los distintos modelos de gobernanza y garantías de control recomiendo en pp. 254 y ss. a BOIX, A., "Los Algoritmos Son Reglamentos: La Necesidad de Extender Las Garantías Propias de Las Normas Reglamentarias a Los Programas Empleados Por La Administración Para La Adopción de Decisiones.", Revista de Derecho Público: Teoría y Método, vol. 1, pp. 223–270. https://doi.org/https://doi.org/10.37417/RPD/vol_1_2020_33.

44. ZUIDERVEEN BORGESIU, F.J., "Strengthening Legal Protection against Discrimination by Algorithms and Artificial Intelligence.", op cit., p. 13.

justifica la necesidad de este análisis y la oportunidad de realizar una crítica constructiva a su propuesta.

A modo de conclusión, se recogen a continuación algunas recomendaciones concretas que se han desarrollado a lo largo del texto:

- Resultaría conveniente que el tercer criterio (“la gravedad de los posibles daños o lesiones causados”) establecido por el artículo 14, para evaluar si una tecnología de IA es o no de alto riesgo, tuviera mayor concreción. Al menos, con la inclusión de los subcriterios contenidos en el Considerando 11 (“la magnitud de la lesión o daño potencial, el número de personas afectadas, el valor total del perjuicio ocasionado y el daño a la sociedad en su conjunto”).
- En el apartado primero del artículo 7 que define la clase de intervención humana requerida para una IA de alto riesgo, sería oportuno que se añadiese a la garantía de una “supervisión humana integral” el criterio cualitativo “y significativa”. Asimismo, es necesario un desarrollo teórico más profundo para que este concepto aporte la seguridad jurídica necesaria.
- La definición de sesgo en el artículo 4 l) resulta muy limitada para comprender lo que se entiende por “sesgo” en el contexto de la IA. Sería recomendable una definición que conjugase la dimensión estadística y ética de los sesgos, basado en el texto de HILDEBRANDT, se propone la siguiente definición: l) «sesgo», toda desviación o representación estadística que reconfigura la distribución de los bienes, servicios, riesgos y oportunidades de una o varias personas físicas o jurídicas.
- Resulta complicado imaginar hasta qué punto las autoridades de control van a poder reforzar el cumplimiento de los principios éticos recogidos por la propuesta sin un adecuado régimen sancionador al que acudir ante eventuales incumplimientos. El cumplimiento normativo debería ser la prioridad ante tecnologías que coincidimos en calificar de alto riesgo para las personas y nuestra sociedad en conjunto.

BIBLIOGRAFÍA

- ALMADA, M., “Human Intervention in Automated Decision-Making: Toward the Construction of Contestable Systems.”, *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*, 2019, pp. 2–11. <https://doi.org/10.1145/3322640.3326699>
- BOIX, A., “Los Algoritmos Son Reglamentos: La Necesidad de Extender Las Garantías Propias de Las Normas Reglamentarias a Los Programas Empleados Por La Administración Para La Adopción de Decisiones.”, *Revista de Derecho Público: Teoría y Método*, vol. 1, pp. 223–270. https://doi.org/10.37417/RPD/vol_1_2020_33
- COBBE, J. y SINGH, J., “Reviewable Automated Decision-Making.”, *Computer Law & Security Review*, vol. 39, 2020. <https://doi.org/10.1016/j.clsr.2020.105475>
- COMISIÓN EUROPEA (CE), “Libro Blanco sobre la inteligencia artificial - un enfoque europeo orientado a la excelencia y la confianza”, Bruselas, COM(2020) 65 final, 19 de febrero de 2020. Disponible en: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_es.pdf

- DANKS, D. y LONDON, A.J., "Algorithmic Bias in Autonomous Systems." Proceedings of the 26th International Joint Conference on Artificial Intelligence, AAAI Press, 2017, pp. 4691–4697. <https://dl.acm.org/doi/10.5555/3171837.3171944>
- DEMETZOU, K., "Data Protection Impact Assessment: A Tool for Accountability and the Unclassified Concept of 'High Risk' in the General Data Protection Regulation.", *Computer Law & Security Review*, vol. 35, núm. 6, 2019. <https://doi.org/10.1016/j.clsr.2019.105342>
- FISCHER, J.E., GREENHALGH, C., JIANG, W., RAMCHURN, S.D., WU, F. y RODDEN, T., "In-the-loop or on-the-loop? Interactional arrangements to support team coordination with a planning agent.", *Concurrency and Computation: Practice and Experience*, 2017, pp. 1-16. <https://doi.org/10.1002/cpe.4082>
- GRUPO DE EXPERTOS DE ALTO NIVEL SOBRE IA (HLEG-AI), "Directrices éticas para una IA fiable", Bruselas, 8 de abril de 2019. Disponible en: <https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1>
- GRUPO DE TRABAJO SOBRE PROTECCIÓN DE DATOS DEL ARTÍCULO 29 (GT29), "Directrices Sobre Decisiones Individuales Automatizadas y Elaboración de Perfiles a Los Efectos Del Reglamento 2016/679.", Bruselas, 2018, pp. 1-37. Disponible en: <https://www.aepd.es/sites/default/files/2019-12/wp251rev01-es.pdf>
- HILDEBRANT, M., "The issue of bias: the framing powers of ML", Draft version, en Marcello Pellillo, Teresa Scantamburlo (eds.), *Machine Learning and Society: Impact, Trust, Transparency*, MIT Press forthcoming 2020
- KISELEVA, A., "AI as a Medical Device: Is It Enough to Ensure Performance Transparency and Accountability?", *European Pharmaceutical Law Review*, vol. 4, núm. 1, 2020, pp. 5-16. <https://doi.org/10.21552/eplr/2020/1/4>
- De LAAT, P.B., "Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?", *Philosophy & Technology*, vol. 31, núm. 4, 2018, pp. 525–41. <https://doi.org/10.1007/s13347-017-0293-z>
- MANN, M. y MATZNER, T., "Challenging Algorithmic Profiling: The Limits of Data Protection and Anti-Discrimination in Responding to Emergent Discrimination.", *Big Data & Society*, vol. 6, núm. 2, 2019, pp. 1–11. <https://doi.org/10.1177%2F2053951719895805>
- MITTELSTADT, B. 2019. "Principles Alone Cannot Guarantee Ethical AI.", *Nature Machine Intelligence*, vol. 1, núm. 11, 2019, pp. 501–507. <https://doi.org/10.1038/s42256-019-0114-4>
- MITTELSTADT, B., "From Individual to Group Privacy in Big Data Analytics.", *Philosophy & Technology*, vol. 30, núm. 4, 2017, pp. 475–94, <https://doi.org/10.1007/s13347-017-0253-7>
- De SIO, F. y Van Den HOVEN, J., "Meaningful Human Control over Autonomous Systems: A Philosophical Account.", *Frontiers in Robotics and AI*, vol. 5, 2018, pp. 1-15. <https://doi.org/10.3389/frobt.2018.00015>
- VEALE, M. y BINNS, R., "Fairer Machine Learning in the Real World: Mitigating Discrimination without Collecting Sensitive Data.", *Big Data & Society*, vol. 4, núm. 2, 2017, pp. 1-17. <https://doi.org/10.1177%2F2053951717743530>
- WACHTER, S. "Affinity Profiling and Discrimination by Association in Online Behavioural Advertising." *Berkeley Technology Law Journal*, vol. 35, núm. 2, 2020, Forthcoming, pp. 1-74. <https://dx.doi.org/10.2139/ssrn.3388639>
- ZUIDERVEEN BORGESIU, F.J., "Strengthening Legal Protection against Discrimination by Algorithms and Artificial Intelligence.", *The International Journal of Human Rights*, vol. 24, núm. 10, 2020, pp. 1–22. <https://doi.org/10.1080/13642987.2020.1743976>