Revista CIDOB d'Afers Internacionals
ISSN: 1133-6595
ISSN: 2013-035X
publicaciones@cidob.org
Barcelona Centre for International Affairs
España

Levitina, Anna
Los seres humanos en la toma de decisiones automatizada en el marco del RGPD y la Ley de IA
Revista CIDOB d'Afers Internacionals, núm. 138, 2024, pp. 121-141
Barcelona Centre for International Affairs
España

DOI: https://doi.org/10.24241/rcai.2024.138.3.121

Disponible en: https://www.redalyc.org/articulo.oa?id=695780434007

Revista CIDOB d'Afers Internacionals
n.º 138, p. 121-144
ISSN:1133-6595 | E-ISSN:2013-035X
DOI: doi.org/10.24241/rcai.2024.138.3.121

Fecha de recepción: 20.03.24
Fecha de aceptación: 05.08.24

Los seres humanos en la toma de decisiones automatizada en el marco del RGPD y la Ley de IA

Humans in automated decision-making under the GDPR and AI Act

Anna Levitina
Abogada especializada en tecnología, protección y privacidad de datos, y sistemas de toma de decisión automatizada (ADS); investigadora predoctoral, Universitat Autònoma de Barcelona (UAB). *1592113@uab.cat. ORCID: https://orcid.org/0009-0004-9855-5334*

Cómo citar este artículo: Levitina, Anna. «Los seres humanos en la toma de decisiones automatizada en el marco del RGPD y la Ley de IA». *Revista CIDOB d'Afers Internacionals*, n.º 138 (diciembre de 2024), p. 121-144. DOI: doi.org/10.24241/rcai.2024.138.3.121

Resumen: La supervisión humana es fundamental para evitar que las máquinas emitan juicios inadecuados sobre las personas a las que se dirigen las decisiones. Aunque el Reglamento General de Protección de Datos (RGPD) y la Ley de Inteligencia Artificial (IA) de la UE abordan esta cuestión, son insuficientes en los aspectos cualitativos y en la integración de supervisores humanos en los marcos de gobernanza. Este artículo examina los requisitos legales para la supervisión humana, analizando cómo estos se entrelazan con las obligaciones de rendición de cuentas de los responsables del despliegue de la toma de decisiones automatizada (ADM) y los derechos individuales. Se aboga por un enfoque más global que no solo incluya la supervisión humana, sino también la evaluación rigurosa y continua de la eficacia del control humano. Sin ello, la supervisión humana puede no proteger adecuadamente el impacto de la ADM en las personas afectadas.

Palabras clave: supervisión humana, Reglamento General de Protección de Datos (RGPD), inteligencia artificial (IA), toma de decisiones automatizadas (ADM), Unión Europea (UE)

Abstract: *Human oversight is a fundamental safeguard against inappropriate judgments made by machines about people who are the targets of these decisions. Although the General Data Protection Regulation (GDPR) and the AI Act address human oversight to some extent, they fall short in addressing qualitative aspects and the integration of human overseers into governance frameworks. This paper examines the legal requirements for human oversight, investigating how these intersect with the accountability obligations of the automated decision-making (ADM) deployers and individual rights. It argues for a more comprehensive approach that not only includes human oversight, but also a continuous and rigorous assessment of the effectiveness of human control. Without that, human oversight may fail to protect adequately and could even worsen the impact on individuals affected by ADM.*

Key words: *human oversight, General Data Protection Regulation (GDPR), artificial intelligence (AI), automated decision-making (ADM), accountability, European Union (EU)*

Ante la creciente presencia de la inteligencia artificial (IA) en nuestras vidas, la cuestión de cómo se gobiernan los sistemas de toma de decisiones automatizadas (ADMS, por sus siglas en inglés), incluidos los sistemas de IA, adquiere una importancia fundamental. El concepto de gobernanza mediante intervención humana, que implica mantener la supervisión y el control humanos sobre los sistemas de IA, es un aspecto clave de este debate (Lazcoz y De Hert, 2022). El Reglamento General de Protección de Datos (RGPD) de la Unión Europea (UE), en vigor desde mayo de 2018, estableció un marco fundacional para la participación humana en la toma de decisiones automatizada (ADM, por sus siglas en inglés), exigiendo una participación humana significativa en estos procesos. Más tarde, el Reglamento de IA de la UE del 13 de junio de 2024, conocido también como la Ley de IA (AIA, por sus siglas en inglés [AI Act]), que reconoce la necesidad de una supervisión humana más exhaustiva cuando se empleen sistemas de IA de alto riesgo, exige además que las personas supervisoras posean la competencia, la formación y la autoridad necesarias, y que se garantice su nivel de conocimientos sobre IA. Sin embargo, a pesar de estos avances, el RGPD y la Ley de IA son insuficientes para abordar los aspectos cualitativos de la supervisión humana y de integrar a las personas supervisoras en los modelos de gobernanza.

Este artículo sostiene que la gobernanza humana obligatoria exige una mayor elaboración. Así, examina, en primer lugar, los requisitos legales de la supervisión humana, analizando cómo estos se entrelazan con las obligaciones de rendición de cuentas de los responsables del despliegue de la ADM y los derechos individuales. A continuación, analiza la forma de integrar eficazmente a las personas supervisoras en los modelos de gobernanza de los responsables del despliegue de la ADM. En este caso, las personas supervisoras se consideran parte integrante de los procesos compuestos de toma de decisiones que afectan a las personas protegidas por la ley, en lugar de partes interesadas independientes. Partiendo del marco propuesto por Lazcoz y De Hert (2022) –que sostiene que la intervención humana debe regirse por mecanismos de gobernanza– y reconociendo la reciente contribución de la Agencia Española de Protección de Datos (AEPD, 2024) –que subraya que la evaluación del grado de intervención debe tener en cuenta la cualificación y diligencia de las personas supervisoras–, este trabajo afirma que las medidas organizativas y las herramientas de evaluación de impacto no solo deben abordar la presentación y el sentido de la intervención humana, sino que también deben incluir criterios de competencia, conocimiento y carácter moral de las personas intervinientes. En el artículo se aboga por una perspectiva holística con respecto a la toma de decisiones automatizada supervisada por humanos (HADM, por sus siglas en inglés), tratando tanto a los agentes algorítmicos como a los humanos como sus

componentes integrales. La gobernanza eficaz de la HADM en su conjunto es esencial para garantizar que los sistemas de ADM subordinados al ser humano beneficien a los intereses de las personas y de la sociedad civil.

La toma de decisiones automatizada (ADM) y los humanos

En los últimos años, la ADM –el proceso de tomar decisiones exclusivamente por medios automatizados sin intervención humana directa– ha reconfigurado la toma de decisiones tradicional, gracias a ventajas como el procesamiento rápido de datos, la escalabilidad y la reducción del error humano. Los sistemas de ADM (ADMS, por sus siglas en inglés) analizan grandes cantidades de datos, detectan patrones y generan resultados, incluidas predicciones y decisiones, por lo que son herramientas valiosas en numerosos sectores, como las finanzas, la sanidad y la justicia penal. Los riesgos asociados a las tecnologías que sustentan la ADM –en particular la IA, capaz de sustituir a los agentes humanos– han impulsado su regulación legislativa. El RGPD, que surgió como una regulación tecnológicamente neutral, instituyó una amplia prohibición de la ADM e introdujo salvaguardias adicionales para los titulares de los datos personales en este ámbito. Recientemente, este marco normativo se ha complementado con la Ley de IA que, si bien se alinea con el RGPD, se centra específicamente en los sistemas de IA (AIS, por sus siglas en inglés) y sus riesgos asociados.

Los ADMS aprenden de datos históricos y pueden contener sesgos intrínsecos que, si no se abordan adecuadamente, pueden conducir a resultados discriminatorios, perpetuando las desigualdades sociales y afectando de manera desproporcionada a las poblaciones vulnerables (Almada, 2019; Rovatsos *et al.*, 2019). Un ejemplo de ello es el ADMS del Ministerio del Interior del Reino Unido para las solicitudes de visado, que se suspendió tras las acusaciones de «racismo arraigado» debido al uso de la nacionalidad como factor de riesgo en la evaluación de los solicitantes. Además, muchos algoritmos avanzados funcionan como «cajas negras», lo que dificulta la comprensión de sus procesos de toma de decisiones (Schmidt *et al.*, 2020). Esta falta de transparencia y explicabilidad plantea cuestiones de rendición de cuentas e impide a las personas cuestionar las decisiones automatizadas, lo que limita su control sobre importantes resultados que podrían cambiar su vida (Edwards y Veale, 2018; Roig, 2020). La trazabilidad limitada y las limitaciones de las auditorías a los ADMS dificultan la identificación y rectificación de errores, así como la evaluación del cumplimiento de

la normativa y la adopción oportuna de medidas correctivas (Berger *et al.*, 2021; Berger *et al.*, 2023). Un ejemplo de ello es el sistema SyRI de los Países Bajos, utilizado para detectar el fraude en la asistencia social y suspendido tras seis años de funcionamiento debido a prácticas discriminatorias basadas en el nivel de ingresos y el origen étnico.

Desde una perspectiva más amplia, la ADM plantea amenazas a los derechos humanos fundamentales, como la protección de datos, la privacidad, la dignidad y autonomía humanas, la ausencia de discriminación y la buena administración. Los ADMS pueden pasar por alto los matices de las circunstancias individuales, reduciendo a las personas a meros datos en algoritmos, o afectar a la autonomía personal, especialmente cuando las decisiones automatizadas son vinculantes y tienen consecuencias a largo alcance. Si se utilizan en el sector público, la ADM puede comprometer la administración pública, privando a las personas de decisiones justas, justificadas e impugnables (Misuraca y Van Noordt, 2020). Por eso, la rendición de cuentas en la ADM, aunque compleja, es fundamental. En caso de incumplimiento de las reglas o normas sustantivas y de procedimiento, debe asignarse la responsabilidad de las acciones (u omisiones) y sus consecuencias, y las personas afectadas deben poder disponer de una compensación efectiva. Sin embargo, existen dudas sobre si los mecanismos de rendición de cuentas existentes son adecuados, dada la implicación de múltiples partes en el desarrollo, despliegue y uso de los ADMS, así como su naturaleza compleja, capacidad de aprendizaje e imprevisibilidad (Wieringa, 2023; Wagner, 2019).

La supervisión humana puede no ser suficiente para mitigar los riesgos asociados a los sistemas de toma de decisiones automatizadas (ADMS), lo que pone de relieve la necesidad de asignar claramente funciones y responsabilidades, así como de incluir a las personas afectadas y a la sociedad en general en la metodología de supervisión.

Investigaciones recientes subrayan la necesidad de comprender claramente lo que esto implica en términos de responsabilidad, reconocimiento de la autoridad y limitación del poder (Novelli *et al.*, 2024). La supervisión humana se cita con frecuencia como mecanismo primordial para garantizar que los ADMS operen dentro de los límites de la ley y respeten los derechos fundamentales. Sin embargo, la creciente preocupación sugiere que la supervisión humana puede no ser suficiente para mitigar los riesgos asociados a los ADMS, lo que pone de relieve la necesidad de asignar claramente funciones y responsabilidades, así como de incluir a las personas afectadas y a la sociedad en general en la metodología de supervisión (Kyriakou y Otterbacher, 2023; Green, 2022; Koulu, 2020).

El RGPD: titulares de los datos y responsables del tratamiento de datos personales

La toma de decisiones automatizadas (ADM) y la intervención humana en el marco del RGPD

El artículo 22 del RGPD constituye el núcleo del marco regulador de la ADM. Según su apartado 1, la ADM hace referencia a aquellas decisiones basadas únicamente en el tratamiento automatizado, incluida la elaboración de perfiles (el análisis automatizado de datos personales para hacer predicciones o tomar decisiones sobre las personas), que produzcan efectos jurídicos o afecten significativamente de modo similar a las personas. Aunque está concebido como un derecho del titular de los datos, dicho apartado funciona como una prohibición general de la ADM, estableciendo que los ADMS deben implicar una revisión humana antes de que se tome la decisión final. Esta disposición está sujeta a las excepciones detalladas en el apartado 2 del artículo 22, que permiten la ADM sin intervención humana previa, ya sea con fines contractuales, con el consentimiento explícito del titular de los datos, o según lo autorice el derecho de la Unión o de los estados miembros. En los casos en que se permita la ADM, el apartado 3 del artículo 22 exige que los titulares de los datos tengan derecho a obtener una intervención humana *a posteriori*, a expresar su punto de vista y a impugnar la decisión.

La interpretación de los términos clave del apartado 1 del artículo 22 ha dado lugar a un debate sobre lo que constituye una «decisión», los criterios para que la decisión se reconozca como exclusivamente automatizada, el alcance de los «efectos jurídicos» y los «efectos significativamente similares», y si el tratamiento de datos no personales entra en el ámbito del artículo 22 (Bygrave, 2020; Binns y Veale, 2021; Mendoza y Bygrave, 2017). Para abordar estas ambigüedades, el Comité Europeo de Protección de Datos (CEPD) y su predecesor, el Grupo de Trabajo del artículo 29 (GT29), publicaron las Directrices sobre Decisiones Individuales Automatizadas y Elaboración de Perfiles («las Directrices»), que contienen explicaciones y ejemplos prácticos para garantizar una interpretación uniforme de las disposiciones del RGPD relacionadas con la ADM (Directrices WP251, 2017). En ellas se subraya la necesidad de que cualquier ADM cuente con una intervención humana *significativa* para evitar que se clasifique como totalmente automatizada, y advierten además contra los intentos de eludir la normativa mediante una participación humana superficial. Para que se considere significativa, la intervención humana debe ir acompañada de la autoridad y la competencia necesarias para modificar la decisión; ya que, aunque en el proceso

intervenga un ser humano, la mera aplicación de perfiles automatizados a las personas sin ninguna influencia sustancial en el resultado sigue entrando en la categoría de la ADM. Las Directrices destacan que los responsables humanos de la toma de decisiones deben tener en cuenta múltiples fuentes de información para mitigar los riesgos asociados a una dependencia excesiva de los resultados automatizados. Estas instrucciones son igualmente pertinentes para las intervenciones humanas realizadas a priori y a posteriori (artículo 22, apartados 1 y 3).

No obstante, debe tenerse en cuenta que el despliegue de agentes humanos no garantiza automáticamente la consideración de los intereses de los titulares de los datos, ni refuerza la protección de sus derechos fundamentales; los responsables del tratamiento de datos pueden ser incapaces de establecer requisitos y controles adecuados para la decisión humana (HDM, por sus siglas en inglés), y las características personales de los responsables humanos pueden no ser las apropiadas para alcanzar estos fines. Además de una posible baja cualificación y falta de conocimientos y experiencia, los agentes humanos pueden mostrar una amplia gama de comportamientos deficientes al interactuar con la ADM, basados en sus propias ideas preconcebidas, como una aversión indebida o una confianza excesiva en los ADMS, y también pueden incorporar su propio sesgo, comprometiendo potencialmente las decisiones que deberían supervisar (Kern *et al.*, 2022; Logg *et al.*, 2019; Burton *et al.*, 2019; Alexander *et al.*, 2018).

Además de una posible baja cualificación y falta de conocimientos y experiencia, los agentes humanos pueden mostrar una amplia gama de comportamientos deficientes al interactuar con la toma de decisiones automatizada (ADM) basados en sus propias ideas preconcebidas.

Los responsables del tratamiento de datos y la calidad de la intervención humana en la ADM

El principio básico del enfoque del RGPD es la rendición de cuentas del responsable del tratamiento de datos personales (RGPD: artículo 5, apartado 2). Consagrada en los artículos 5 y 24, dicha responsabilidad obliga a los responsables del tratamiento de datos a demostrar su cumplimiento mediante sólidas medidas técnicas y organizativas. El artículo 25, apartado 1, obliga además a integrar las medidas de protección de datos desde el inicio de las actividades de tratamiento, y el artículo 35 exige evaluaciones de impacto en la protección de datos (EIPD) cuando estén presentes ADM, lo que permite una evaluación sistemática de los posibles riesgos y contramedidas (Kaminski y Malgieri, 2021: 140). Además, el artículo 37 exige el nombramiento de

delegados de protección de datos (DPD) en determinadas circunstancias para supervisar y controlar el cumplimiento por parte de los responsables del tratamiento. El DPD puede ser instrumental para garantizar la significatividad de la HDM (Sartor y Lagioia, 2020).

En el marco del RGPD, los responsables del tratamiento de datos tienen responsabilidades específicas en materia de ADM. Según lo estipulado en el artículo 22 del RGPD, dichos responsables deben o bien implementar un proceso significativo de revisión humana de los resultados automatizados, destinado a proteger a los titulares de los datos ante la ADM (para cumplir la prohibición del artículo 22, apartado 1), o bien ofrecer una revisión humana de una decisión automatizada ya adoptada (según lo permitido por el artículo 22, apartado 2) a petición del titular de los datos (para cumplir el artículo 22, apartado 3). Los procedimientos y procesos introducidos por los responsables del tratamiento para garantizar la intervención humana requerida pueden variar, siempre que garanticen que la intervención sea significativa (Almada, 2021; Wagner, 2019).

Lazcoz y De Hert (2022) sostienen que, además de ser un requisito de cumplimiento, la intervención humana también es fundamental para garantizar la rendición de cuentas: los responsables del tratamiento de datos son, en última instancia, responsables de las decisiones automatizadas. Sin embargo, aunque el RGPD los responsabiliza de la aplicación de las medidas pertinentes, incluida una intervención humana significativa, no impone requisitos de calidad para la HDM, la cual también forma parte del proceso de la HADM en su conjunto. El principal objetivo del RGPD es proteger a los titulares de los datos: la inclusión de la HDM debidamente considerada, documentada y cualificada en combinación con la ADM en el marco de la rendición de cuentas es esencial para reforzar la protección de los titulares de los datos y la buena gobernanza.

La evolución de la ADM nos impulsa a encontrar un delicado equilibrio entre el potencial innovador de la tecnología y la salvaguardia de los derechos fundamentales. Los responsables humanos de la toma de decisiones se perfilan como agentes fundamentales en este equilibrio. Sin embargo, la participación humana no inmuniza ni a la ADM ni a la HADM frente a problemas como la parcialidad, la discriminación o la falta de explicación o justificación (Kern *et al.*, 2022; Yeung, 2018). Dada la necesidad de que los agentes humanos participen de manera significativa en el proceso de toma de decisiones o intervengan adecuadamente a petición del titular de los datos, y teniendo en cuenta su papel a la hora de facilitar la rendición de cuentas de los responsables del tratamiento de datos, es prudente que dichos responsables integren requisitos de calidad para la HDM en diversos aspectos del marco de rendición de cuentas relacionado con la ADM.

En primer lugar, reconociendo que los responsables del tratamiento de datos y los responsables humanos de la toma de decisiones no son equivalentes, los primeros podrían aplicar, como parte de sus medidas organizativas, ciertas políticas y procedimientos internos dirigidos a los responsables humanos de la toma de decisiones que actúan como agentes de los responsables del tratamiento de datos. Las medidas, además de garantizar la presencia de un revisor humano en la ADM, abarcarían las cualificaciones y competencias de los revisores, exigirían aptitudes específicas y normas de rendimiento, e incluirían criterios y procedimientos de evaluación relativos a los responsables humanos de la toma de decisiones. Incluir a estos últimos de este modo es obligatorio tanto si los procesos de toma de decisiones suponen la ADM como en caso contrario: bien para justificar la ausencia de la ADM, bien para establecer los mecanismos y el contenido de la HDM previa solicitud. Además, dado que los principios y requisitos generales del RGPD se aplican al tratamiento de datos parcialmente automatizados (en contraste con las decisiones totalmente automatizadas cubiertas por el artículo 22), dicha inclusión de agentes humanos mejora el cumplimiento general y la rendición de cuentas de los responsables del tratamiento de datos a través del marco de la protección de la privacidad desde el diseño.

En segundo lugar, el RGPD introduce la obligación de que los responsables del tratamiento de datos realicen EIPD cuando esté implicada la ADM (Directrices WP251, 2017). Estos análisis van más allá de la evaluación de riesgos: incluyen una visión global de las medidas destinadas a mitigar los riesgos y demostrar el cumplimiento del RGPD, teniendo en cuenta los derechos e intereses de los titulares de los datos y otras personas involucradas. Si bien es probable que una EIPD describa los riesgos para los titulares de los datos y las medidas de salvaguardia existentes, incluida la intervención humana prevista, sostenemos que, como en el caso de las medidas organizativas, también debe incluir una evaluación (requisitos pertinentes y revisión del rendimiento) de las propias personas que intervienen.

Por último, el artículo 37 del RGPD introduce una función humana más en el ámbito de los responsables del tratamiento de datos: un delegado de protección de datos (DPD). Como supervisores independientes y con competencias consultivas del cumplimiento por parte de los responsables del tratamiento de datos, los DPD podrían contribuir a la protección de los titulares de los datos reforzando el marco de rendición de cuentas y añadiendo una perspectiva independiente a la gobernanza de la ADM, ofreciendo orientación al responsable del tratamiento sobre las EIPD y supervisando posteriormente su aplicación, además de ocuparse de las consultas de los titulares de los datos relacionadas con el tratamiento de sus datos personales y sus derechos.

Aunque la existencia de la ADM no requiere necesariamente el nombramiento de un DPD, esta figura puede servir como doble punto de control dentro del marco normativo del RGPD y de los procesos de ADM. Su conocimiento de las operaciones del responsable del tratamiento de datos y de las prácticas de tratamiento relacionadas con la ADM, les capacita para facilitar una comunicación y cooperación eficaces con los organismos reguladores. En esencia, los DPD podrían tender un puente entre la gobernanza interna y el control regulatorio externo (Comité Europeo de Protección de Datos, 2024), mejorando la transparencia y la rendición de cuentas en el ámbito de la ADM. Por otra parte, como supervisores internos de los responsables del tratamiento de datos, los DPD están en condiciones de garantizar que dichos responsables incorporen requisitos de calidad y procedimientos de control en sus procesos operativos para la intervención humana en la ADM, promoviendo así una mejor gobernanza y garantizando que los agentes humanos interactúen con los ADMS de manera que se beneficie a los titulares de los datos (Roig, 2017). Sin embargo, incluso a este nivel de experiencia, debe tenerse en cuenta que las ventajas de un DPD están condicionadas a su competencia personal y a su comprensión de las cambiantes tecnologías con las que trabajan (ibídem).

La inclusión de la decisión humana (HDM) debidamente considerada, documentada y cualificada en combinación con la ADM en el marco de la rendición de cuentas es esencial para reforzar la protección de los titulares de los datos y la buena gobernanza.

Los titulares de los datos y la toma de decisiones humana

En virtud del RGPD, los titulares de los datos tienen derecho a una revisión humana por defecto (cuando la ADM está prohibida según el artículo 22, apartado 1) y, en caso de estar permitida, a petición suya (según el artículo 22, apartado 3). La revisión humana debe ser significativa, implicar la consideración de otras fuentes de información y ser llevada a cabo por un revisor humano autorizado (Directrices WP251, 2017: 21). Además, cuando se permite la ADM, los titulares de los datos tienen derecho a expresar sus opiniones e impugnar las decisiones automatizadas (artículo 22, apartado 3). Aunque dicho apartado no aclara explícitamente si la expresión de puntos de vista o la impugnación de decisiones deben dirigirse a un revisor automatizado o a un revisor humano, el árbitro último en materia de la ADM, lógicamente, no debería ser otra capa de ADM. En otras palabras, se requiere un revisor humano en todas las circunstancias. Los titulares de los datos, sin embargo, tienen que confiar en la HDM proporcionada, sin poder influir en su calidad.

Los titulares de los datos también tienen derecho a ser informados sobre la ADM, tal como se establece en el artículo 13, apartado 2, letra f, y en el artículo 14, apartado 2, letra g. Este derecho implica que se les informe de la existencia de ADM en las prácticas de procesamiento del responsable del tratamiento de datos y que se les proporcione información significativa sobre la lógica que subyace a los ADMS, así como sobre su importancia y posibles implicaciones para el titular de los datos. Aunque las Directrices recomiendan que se informe a los titulares de los datos incluso si las decisiones automatizadas no entran en el ámbito de aplicación del artículo 22, apartado 1, los responsables del tratamiento de datos podrían optar por no proporcionarles información, aprovechando una interpretación restrictiva de los requisitos del RGPD que consideraría que la ADM acompañada de HDM queda fuera del ámbito de aplicación del artículo 22, apartado 1 y, por lo tanto, fuera del requisito de informar a los titulares de los datos de la existencia de ADM (Directrices WP251, 2017: 25). En tales circunstancias, la detección de ADM podría, como mínimo, resultar difícil. Sin información sobre la ADM o la forma en que se aplica, el incumplimiento será efectivamente imposible de detectar, lo que impedirá que el titular de los datos ejerza sus derechos y obstaculizará las medidas para garantizar su aplicación (Sivan-Sevilla, 2024; Lynskey, 2023).

Incluso cuando los titulares de los datos son conscientes de la aplicación de ADM a sus datos personales, resulta problemático determinar qué constituye exactamente la información *significativa* que deben facilitar los responsables del tratamiento de datos. La cuestión de la explicabilidad de la ADM ha sido ampliamente debatida por investigadores e instituciones de la UE (Bauer *et al.*, 2021; Cobbe *et al.*, 2021; Malgieri, 2021; Selbst y Powles, 2017). A este respecto, nos limitamos a señalar que para que los responsables del tratamiento de datos proporcionen información significativa, deben poseer dicha información en primer lugar, lo que no es necesariamente el caso. Puede ocurrir que dichos responsables del tratamiento de datos y sus supervisores humanos de ADMS carezcan de la capacidad necesaria para comprender o acceder a la información sobre las operaciones y los procesos de toma de decisiones de los ADMS. En consecuencia, podrían ser incapaces de proporcionar a los titulares de los datos la información requerida (Grant *et al.*, 2023).

Así pues, hay múltiples factores en juego que pueden exponer a los titulares de los datos a violaciones de sus derechos fundamentales: la responsabilidad de garantizar una intervención humana significativa recae en los responsables del tratamiento de datos; existe la posibilidad de que las decisiones queden descalificadas por el artículo 22, apartado 1, con la correspondiente falta de suministro de información sobre la ADM, y puede haber obstáculos en la aplicación. También existe la posibilidad de que los responsables del tratamiento de datos den

prioridad al cumplimiento de la normativa de protección de datos en lugar de priorizar la auténtica protección de los datos y de sus titulares, lo que podría dar lugar a un «lavado de cumplimiento» por su parte, especialmente en el contexto de la ADM y de la necesaria intervención humana. Nuestra opinión es que, dadas estas dificultades potenciales, sería más lógico abordar la toma de decisiones automatizada supervisada por humanos (HADM) híbrida, informando a los titulares de los datos de la existencia de ADM en todas las circunstancias, así como aumentando los requisitos legales actuales proporcionando información sobre el alcance de la HDM y los fundamentos de la interacción entre la ADM y la HDM presente en una situación determinada. Comunicar a los titulares de los datos la existencia de ADM es crucial para garantizar la transparencia, pudiendo este énfasis adicional en la HDM reforzar la precisión y la fiabilidad al evidenciar a los titulares de los datos que la ADM y la HDM son dos facetas del mismo proceso de toma de decisiones. Esto se podría facilitar a través de las vías existentes disponibles en virtud del RGPD, como el derecho de acceso a los datos y las evaluaciones de impacto en la protección de datos (EIPD). Sin embargo, es fundamental reconocer que estas vías tienen limitaciones inherentes que deben reconocerse y abordarse para garantizar una transparencia y una rendición de cuentas efectivas.

Asimismo, si bien los responsables del tratamiento de datos deben proporcionar la información relativa a la ADM en el momento en que se recogen los datos personales o en torno a ese momento (artículos 13 y 14), los titulares de los datos conservan el derecho a acceder (solicitar) información equivalente en cualquier momento (artículo 15). La información facilitada en virtud del artículo 15 puede tener que actualizarse o adaptarse para reflejar las operaciones de tratamiento específicas del titular de los datos solicitante (Custers y Heijne, 2022). Esta distinción entre la obligación de facilitar información y el derecho a acceder a ella tiene por objeto permitir a los titulares de los datos verificar la exactitud de los datos y la legalidad del tratamiento. Los titulares de los datos que ejercen su derecho de acceso pueden obtener información adicional sobre la ADM, también sobre datos inferidos o derivados, como resultados algorítmicos o personalizados, y detalles sobre la justificación de las decisiones tomadas sobre su persona, en lugar de sobre la lógica de la ADM en general (ibídem: 5). Sin embargo, sigue pendiente la cuestión de si proporcionar a un titular de los datos información concreta sobre una decisión específica contribuye realmente a evaluar la ecuanimidad, imparcialidad y legitimidad de dicha decisión. Dado que es poco probable que el titular de los datos tenga conocimiento de decisiones análogas relativas a otras personas para identificar los factores que provocan disparidades en la toma de decisiones o validar su coherencia, esta cuestión sigue siendo discutible (Dreyer y Schulz, 2019). A primera vista, el derecho de acceso

refuerza la protección de los datos individuales, al tiempo que fomenta la búsqueda de la equidad social y el interés público cuando se ejerce colectivamente en la dimensión de la sociedad civil, aunque hasta el momento ha habido poco uso de esta posibilidad (Mahieu y Ausloos, 2020a).

Las EIPD exigidas por el RGPD ofrecen una vía adicional para que los titulares de los datos participen en la gobernanza de la ADM: los responsables del tratamiento de datos pueden solicitar la opinión de los titulares de los datos (y de sus representantes) sobre el procesamiento previsto. Sin embargo, los titulares de los datos pueden quedar al margen si los responsables del tratamiento consideran que no es apropiado recabar su opinión, en particular si están en juego intereses comerciales o públicos, o la seguridad de las operaciones de tratamiento de datos (RGPD: artículo 35, apartado 9). Otro problema asociado a este mecanismo es que las consultas pueden convertirse en meras formalidades debido a factores como las opciones de diseño y contenido de las EIPD, las capacidades de las personas y su disposición a desarrollar un conocimiento detallado del tema o a participar activamente en las EIPD (Christofi *et al.*, 2022). A la luz de lo anterior, la designación de un representante de los titulares de los datos –como una organización sin ánimo de lucro o una asociación especializada en la protección de los titulares de los datos– podría ser una solución más eficaz, ya que defendería los intereses de los titulares de los datos al modo de un grupo de protección de los consumidores. Aunque ya existen varias entidades de defensa, su capacidad y recursos siguen siendo limitados (Mahieu y Ausloos, 2020b).

Una vez más, en situaciones en las que las EIPD encuentran dificultades para lograr la participación efectiva de los titulares de los datos o garantizar la adecuada representación de sus intereses, un DPD puede ser de gran utilidad, al garantizar el cumplimiento de los requisitos del RGPD y defender los derechos de los titulares de los datos.

La Ley de IA: proveedores, responsables del despliegue y personas afectadas

Supervisión humana según la Ley de IA e intervención humana según el RGPD

La Ley de IA (AIA, por sus siglas en inglés) introduce medidas reguladoras adicionales relativas a la ADM y se entrecruza con el RGPD en varios frentes. Mientras que el RGPD aspiraba inicialmente a ser un marco independiente de

la tecnología, la Ley de IA se dirige explícitamente a los sistemas de IA (AIS, por sus siglas en inglés), y su objetivo general es establecer un marco jurídico en el que la IA priorice de forma sistemática a los seres humanos (AIA: Considerando 6). Dicha Ley designa la agencia humana (al servicio de las personas) y la supervisión humana (supervisión adecuada por parte de seres humanos) como principios rectores primordiales para el desarrollo y el uso de la IA en todos los niveles de riesgo (AIA: Considerando 27). Asimismo, también contempla los requisitos y obligaciones de supervisión humana en relación con los AIS de alto riesgo –aquellos que plantean un riesgo significativo de daño para la salud, la seguridad o los derechos fundamentales de las personas físicas y que, por lo tanto, también podrían entrar en el ámbito de aplicación del artículo 22 del RGPD– en la medida en que dichos sistemas produzcan decisiones con efectos jurídicos o significativamente similares sobre los titulares de los datos. Dado que la Ley de IA se aplicará simultáneamente con el RGPD, estos requisitos para los AIS de alto riesgo, junto con las obligaciones de los proveedores y los responsables de su despliegue, contribuirán a la protección de los ciudadanos en el ámbito de la ADM, así como a la rendición de cuentas de los responsables del tratamiento en el marco del RGPD (AIA: artículo 2, apartado 7).

La noción de supervisión humana en el marco de la Ley de IA parece abarcar un ámbito más amplio y, al mismo tiempo, estar definida con menos precisión que la idea de intervención humana esbozada en el RGPD (Lazcoz y De Hert, 2022: 12). Mientras que el RGPD pretende separar los objetivos de decisión de la ADM mediante el uso de criterios humanos significativos, la Ley de IA exige que el diseño de los AIS incorpore herramientas de interacción humano-máquina que permitan a un humano supervisar los sistemas adecuadamente en relación con los riesgos asociados (AIA: artículo 14). La intervención humana con arreglo al artículo 22 del RGPD podría ser un ejemplo de dicha supervisión humana. El artículo 14 de la Ley de IA exige que los AIS de alto riesgo estén equipados con herramientas adecuadas de interfaz humano-máquina que garanticen la supervisión por parte de personas físicas. El objetivo primordial es la mitigación de los riesgos inherentes a los AIS, incluidos los que afectan a los derechos humanos fundamentales; además, la supervisión debe tener en cuenta los riesgos específicos, el nivel de autonomía y el contexto del AIS (AIA: artículo 14, apartados 2 y 3). Los proveedores tienen la responsabilidad de dotar a las personas encargadas de la supervisión humana de los medios necesarios para comprender las capacidades y limitaciones de los AIS de alto riesgo que están supervisando, interpretar sus resultados, intervenir en su funcionamiento, decidir si utilizarlo (o no) y ser conscientes de la tendencia a confiar ciegamente o en exceso en sus resultados (sesgo de automatización) (AIA: artículo 14, apartado 4).

Responsables del despliegue de sistemas de IA: supervisión humana y las personas supervisoras

Los agentes humanos intervienen a lo largo de todo el ciclo operativo de un AIS, desde la decisión inicial de desplegar el sistema, pasando por su fase de uso, hasta la aplicación de sus resultados en casos concretos. El grado de implicación humana varía, siendo muy comunes enfoques como «human-in-the-loop» (implicación humana activa) y «human-on-the-loop» (supervisión por humanos). Independientemente del grado de autonomía de un AIS, la presencia humana, aunque sea en distintos grados, sigue siendo parte integrante de su funcionamiento. Investigaciones recientes abogan por considerar las interacciones entre humanos y algoritmos como una forma de actuación colaborativa, en lugar de tratar las funciones humanas y automáticas de forma aislada (Tsamados *et al.*, 2024; Green, 2022). Además de servir como mecanismo de control de primera línea, la supervisión por humanos reglamentaria constituye un componente integral del nexo de la toma de decisiones automatizada supervisada por humanos (HADM). Sin embargo, la eficacia de dicha supervisión depende de las aptitudes y la motivación de las personas supervisoras.

El artículo 26 de la Ley de IA exige a los responsables del despliegue que apliquen medidas de supervisión humana en consonancia con las instrucciones facilitadas por el proveedor y garanticen que las personas asignadas para llevar a cabo la supervisión humana tengan la competencia, la autoridad, el apoyo y la formación necesarias para supervisar eficazmente los AIS. Además, los responsables del despliegue deben adoptar medidas proactivas para cultivar un nivel suficiente de conocimientos de IA entre el personal implicado en el funcionamiento y la utilización de los AIS (AIA: artículo 4). Dichas medidas deben adaptarse a los conocimientos técnicos, la experiencia, la educación y la formación de cada persona, así como a los contextos específicos en los que se despliegan los AIS. Las estipulaciones de la Ley de IA relativas a la alfabetización y la formación necesaria en materia de IA adquieren una gran importancia, aunque su alcance sigue resultando más bien limitado.

Una supervisión humana significativa y cualificada, reconocida como un componente vital para salvaguardar los intereses de las personas y la sociedad, debería integrarse en el modelo de gobernanza de los responsables del despliegue cuando empleen la HADM. Para ello, habría que establecer normas de calidad claras para las personas supervisoras, que incluyeran tanto los conocimientos técnicos como las cualidades personales pertinentes para su función en el contexto operativo de la persona responsable del despliegue (Tsamados *et al.*, 2024; Laux, 2023). Asimismo, dicha integración exigiría una evaluación continua de la actuación de las personas responsables de la toma de decisiones en la práctica, para garantizar que

aplican sus conocimientos y habilidades de forma eficaz, que no incurren en el doble peligro de la excesiva confianza o la reticencia a la hora de interactuar con los ADMS, y que no comprometen el rendimiento general de la HADM.

Mediante el establecimiento de criterios de competencia, conocimientos y carácter moral, así como la integración de estos criterios con los requisitos legales existentes, los modelos de gobernanza de los responsables del despliegue se alinearían con el proclamado enfoque integral y consciente del contexto de la gobernanza de la IA, contribuyendo en última instancia a la protección de las personas afectadas y de la sociedad.

Responsables del despliegue de sistemas de IA: evaluación del impacto sobre los derechos fundamentales y gobernanza de la HADM

Anteriormente, hemos hablado de las evaluaciones de impacto en la protección de datos (EIPD) como una herramienta prometedora para las personas responsables del tratamiento de datos encargadas de gobernar la HADM. Ahora exploraremos cómo puede mejorarse el modelo de gobernanza de este tipo de toma de decisiones en el marco de la Ley de IA, sugiriendo formas de incorporar personas supervisoras a dicho modelo.

La Ley de IA obliga a las personas responsables del despliegue de determinados AIS de alto riesgo a elaborar evaluaciones de impacto sobre los derechos fundamentales (EIDF) en las que se detallen el contexto operativo del AIS, los riesgos potenciales y las medidas de salvaguardia, incluida la supervisión humana, que la persona responsable del despliegue aplicará como elemento del marco de mitigación de riesgos (AIA: artículo 27). En el contexto de la ADM, las EIPD y las EIDF deben elaborarse conjuntamente, y pueden ser objeto de escrutinio público cuando las personas responsables del despliegue estén obligadas a publicar sus EIDF (AIA: artículo 27, apartado 4). Juntas, estas evaluaciones podrían constituir una sólida herramienta de salvaguardia para garantizar la transparencia y la calidad de la HADM (Mantelero, 2022).

Dado que la Ley de IA impone requisitos más estrictos a las personas encargadas de la supervisión humana, incluida su alfabetización en IA (competencia, conocimientos y aptitudes), las EIDF podrían servir de marco para una evaluación más exhaustiva y global que tenga en cuenta las capacidades y competencias de los agentes humanos, así como su capacidad para funcionar eficazmente junto a los AIS como parte integrante de la HADM. Las personas supervisoras no solo deben vigilar el funcionamiento de los AIS, sino también mitigar los riesgos asociados a la propia intervención humana, como los sesgos,

los errores y la excesiva confianza en los AIS. Sin embargo, las disposiciones de la Ley de IA se centran principalmente en la evaluación de los riesgos de los AIS y dejan de lado la intrincada dinámica de la HADM, sin garantizar su calidad. Las personas responsables del despliegue necesitan directrices más claras para evaluar la competencia y preparación de sus supervisores humanos.

Este objetivo puede alcanzarse por varias vías. Las personas responsables del despliegue deben demostrar su capacidad de rendición de cuentas elaborando políticas internas que describan explícitamente las cualificaciones, competencias y responsabilidades de las personas supervisoras, así como la forma en que supervisarán los ADMS, intervendrán e informarán de sus conclusiones y contribuirán a la HADM (Crootof *et al.*, 2023). Las EIDF podrían ayudar a establecer un marco adecuado de gobernanza para la HADM que defina claramente las funciones, responsabilidades y normas necesarias tanto para la HDM como para la ADM, mejorando la transparencia y la rendición de cuentas. Las EIDF podrían exigir descripciones detalladas de los puestos de trabajo y evaluaciones periódicas de las personas supervisoras. Estas evaluaciones garantizarían que las personas responsables de la toma de decisiones no solo estén bien versadas en los aspectos técnicos y operativos de los AIS, sino que también sean capaces de evaluar críticamente las implicaciones de la dinámica y los resultados de la HADM (Sterz *et al.*, 2024; Enarsson *et al.*, 2021). Las EIDF podrían enmarcar la formación continua de las personas supervisoras para mantener los conocimientos técnicos pertinentes y el conocimiento de las normas jurídicas y éticas relacionadas con la ADM.

Las evaluaciones de impacto sobre los derechos fundamentales (EIDF) podrían servir de marco para una evaluación más exhaustiva y global que tenga en cuenta las capacidades y competencias de los agentes humanos, así como su capacidad para funcionar eficazmente junto a los sistemas de IA (AIS).

La gobernanza de la HADM debe incluir mecanismos para el seguimiento y la auditoría de la HDM. Las personas supervisoras, habilitadas para intervenir o anular las decisiones automatizadas, deben ser capaces de explicar el fundamento y las consecuencias previstas de sus acciones, así como de reconocer y mitigar sus propios prejuicios y actuar con integridad, prudencia, imparcialidad, sólido juicio moral y benevolencia. Las EIDF podrían evaluar y validar continuamente las medidas de gobernanza de la HADM, así como el modo en que las personas responsables de la toma de decisiones interactúan con los ADMS, auditando la calidad del rendimiento de la HDM. Además, las EIDF podrían incorporar disposiciones para recabar segundas opiniones, incluso del DPD, a fin de añadir un nivel adicional de escrutinio. Al imponer estas medidas, las EIDF podrían garantizar que las personas supervisoras se rijan por las mismas normas rigurosas que los ADMS que supervisan.

Por otra parte, las personas responsables del despliegue podrían establecer mecanismos de retroalimentación que permitan a las personas supervisoras informar de problemas o inquietudes, identificando posibles áreas de mejora en el ADMS o en el marco de gobernanza. El establecimiento de mecanismos de protección de las personas denunciantes y de salvaguardias contra las represalias podría ayudar a los trabajadores que cuestionan la ADM o denuncian fallos del sistema, garantizando que no se vean disuadidos por el temor a la inseguridad laboral o a las repercusiones por parte de sus superiores.

Personas afectadas y sus derechos

Aunque la Ley de IA no parece mejorar el potencial de los mecanismos de recurso colectivo, que era una limitación del RGPD comentada anteriormente, sí incluye salvaguardias adicionales para las personas. De conformidad con el artículo 26, apartado 11, de la Ley de IA, las personas físicas sujetas a la utilización de AIS de alto riesgo que adopten decisiones o presten asistencia en las mismas deberán ser informadas de que están sujetas a la utilización de dichos AIS. Aunque la Ley de IA solo exige la notificación del uso de este tipo de AIS, y no el suministro de información adicional sobre sus implicaciones para las personas, como requieren los artículos 13 y 14 del RGPD, su redacción se extiende más allá del alcance del derecho a ser informado en virtud del RGPD, exigiendo explícitamente la comunicación del uso de AIS que ayudan a tomar decisiones (es decir, incluso si existe la HDM). También amplía el abanico de personas a las que se concede el derecho a ser informadas, ya que no distingue entre datos personales y no personales y no se limita específicamente a los titulares de los datos.

No obstante, esto puede verse socavado por la exclusión de los AIS de la categoría de alto riesgo si no tienen un impacto *material* en los resultados de la toma de decisiones y no plantean riesgos significativos para la salud, la seguridad o los derechos fundamentales de las personas (tal como se indica en el artículo 6, apartado 3, de la AIA). El calificativo de «materialidad» resulta subjetivo, lo que podría permitir eludir los requisitos establecidos para los AIS de alto riesgo. La Ley de IA establece que las personas sujetas al uso de sistemas de IA de alto riesgo deben tener «el derecho de obtener de la persona responsable del despliegue explicaciones claras y significativas sobre el papel del sistema de IA en el procedimiento de toma de decisiones y los principales elementos de la decisión adoptada». Esta disposición ofrece la posibilidad de que, al contar con dicha información, las personas puedan potencialmente analizar y reconstruir las decisiones, lo que permitiría descubrir el papel de la HDM. Sin embargo, este proceso requeriría una disposición para llevar a cabo dicho análisis, algo

que podría interesar más a auditores, organizaciones de protección de datos o autoridades de supervisión que a los individuos en sí.

La Ley de IA intenta garantizar que las personas que supervisan los AIS de alto riesgo en nombre de las personas responsables del despliegue puedan interpretar correctamente sus resultados. La intención de este requisito de interpretabilidad es garantizar que las personas responsables del despliegue sean capaces de proporcionar la explicación obligatoria sobre la decisión automatizada, pero aún está por ver cómo se garantiza una interpretación «correcta» y qué implica una «explicación clara y significativa».

Conclusión

En general, se considera que el control humano es esencial para mitigar los riesgos asociados a los sistemas automatizados de toma de decisiones (ADMS). Al respecto, este artículo ha abordado críticamente el concepto de control humano, cuestionando la suposición de que dicho control es por sí solo positivo sin un examen riguroso de su calidad. Así, abogamos por una perspectiva holística de la toma de decisiones automatizada con intervención humana (HADM), tratando a los agentes algorítmicos y humanos como componentes integrales de la HADM para garantizar que los ADMS presten el mejor servicio a las personas y a la sociedad civil.

Nuestros resultados ponen de relieve que, si bien el RGPD y la Ley de IA ofrecen marcos fundamentales para integrar la supervisión humana en los ADMS, ambos son insuficientes a la hora de abordar las dimensiones cualitativas de dicha supervisión humana. El RGPD exige la intervención humana en determinados procesos de ADM, pero carece de requisitos específicos en cuanto a la cualificación o formación de las personas supervisoras. Del mismo modo, la Ley de IA reconoce la necesidad de que las personas que interactúan con los ADMS posean conocimientos de IA, pero no detalla directrices sobre las normas de calidad y la evaluación continua de su rendimiento. Para salvar esta distancia, es necesaria una estrategia de gobernanza global, que debe incluir revisiones periódicas del rendimiento, mecanismos de retroalimentación continua y desarrollo profesional permanente para las personas supervisoras. Las evaluaciones deben abordar no solo la exactitud e imparcialidad de las decisiones, sino también la eficacia de la gestión de riesgos y el cumplimiento de las normas éticas. Los mecanismos de retroalimentación, incluidos los canales anónimos, son esenciales para crear una cultura de transparencia y mejora continua. Además, la formación continua ayudaría a las personas supervisoras a mantenerse informadas y capacitadas en el funcionamiento de los ADMS.

Los instrumentos reguladores, como las evaluaciones de impacto en la protección de datos (EIPD) y las evaluaciones de impacto sobre los derechos fundamentales (EIDF), deben aprovecharse más eficazmente para incluir requisitos y evaluaciones tanto de la ADM como de la HDM. Este doble enfoque garantiza un planteamiento holístico de la gestión de riesgos, en el que el elemento humano de la HADM está igualmente sujeto a una supervisión y evaluación continuas para mitigar posibles sesgos y preservar la calidad de la toma de decisiones. Ello se ajusta al concepto de gobernanza adaptativa, que postula que los mecanismos de supervisión deben evolucionar en respuesta a la variación de las condiciones y a la aparición de nueva información. Sin embargo, debemos reconocer varias limitaciones. La aplicación de una estrategia de este tipo puede plantear problemas prácticos, como limitaciones organizativas, económicas y de recursos humanos. La eficacia de los mecanismos de retroalimentación depende de la disposición de las personas supervisoras a comprometerse de forma abierta, lo que puede verse influido por la cultura organizativa y las dinámicas de poder. Además, sin mandatos legales para la adopción de medidas específicas, las personas responsables del despliegue de la HADM podrían limitarse a cumplir las normas mínimas de conformidad, descuidando las salvaguardias complementarias.

Referencias bibliográficas

AEPD-Agencia Española de Protección de Datos. «Evaluating human intervention in automated decisions», (4 de marzo de 2024) (en línea) [Fecha de consulta: 01.09.2024] https://www.aepd.es/en/press-and-communications/blog/evaluating-human-intervention-in-automated-decisions

Alexander, Veronika: Blinder, Collin y Zak, Paul J. «Why trust an algorithm? Performance, cognition, and neurophysiology». *Computers in Human Behavior*, vol. 89, (2018), p. 279-288. DOI: https://doi.org/10.1016/j.chb.2018.07.026

Almada, Marco. «Human Intervention in Automated Decision-Making». *ICAIL '19: Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*, (2019), p. 2-11. DOI: https://doi.org/10.1145/3322640.3326699

Almada, Marco. «Automated Decision-Making as a Data Protection Issue». *SSRN*, (2021), p. 1-23 (en línea) [Fecha de consulta: 01.09.2024] https://dx.doi.org/10.2139/ssrn.3817472

Bauer, Kevin; Hinz, Oliver; van der Aalst, Wil y Weinhardt, Christof. «Expl(AI)n It to Me – Explainable AI and Information Systems Research». *Business and Information Systems Engineering*, vol. 63, (2021), p. 79-82. DOI: https://doi.org/10.1007/s12599-021-00683-2

Berger, Armin; Hillebrand, Lars; Leonhard, David; Deußer, Tobias; Feliz de Oliveira, Thiago B. y Dilmaghani, Tim. «Towards Automated Regulatory Compliance Verification in Financial Auditing with Large Language Models». *2023 IEEE International Conference on Big Data (BigData)*, (15-18 de diciembre de 2023), p. 4.626-4.635. DOI: https://doi.org/10.1109/BigData59044.2023.10386518

Berger, Benedikt; Adam, Martin; Rühr, Alexander y Benlian, Alexander. «Watch Me Improve: Algorithm Aversion and Demonstrating the Ability to Learn». *Business and Information Systems Engineering*, vol. 63, (2021), p. 55-68. DOI: https://doi.org/10.1007/s12599-020-00678-5

Binns, Reuben y Veale, Michael. «Is that your final decision? Multi-stage profiling, selective effects, and Article 22 of the GDPR». *International Data Privacy Law*, vol. 11, n.º 4 (2021), p. 319-332. DOI: https://doi.org/10.1093/idpl/ipab020

Burton, Jason W.; Stein, Mari-Klara y Jensen, Tina Blegind. «A systematic review of algorithm aversion in augmented decision making». *Journal of Behavioral Decision Making*, vol. 33, n.º 2 (2019), p. 220-239. DOI: https://doi.org/10.1002/bdm.2155

Bygrave, Lee A. «Article 22: Automated individual decision-making, including profiling», en Kuner, Christopher; Bygrave, Lee A.; Docksey, Christopher y Dreachsler, Laura (eds.) *The EU General Data Protection Regulation (GDPR) – A Commentary*. Oxford: Oxford University Press, 2020, p. 531.

Christofi, Athena; Breuer, Jonas; Wauters, Ellen; Valcke, Peggy y Pierson, Jo. «Data protection, control and participation beyond consent - Seeking the views of data subjects in data protection impact assessments», en Kosta, Eleni; Leenes, Roland y Kamara, Irene (eds.) *Research Handbook on EU Data Protection Law*. Cheltenham: Edward Elgar, 2022, p. 503-529.

Cobbe, Jennifer, Seng Ah Lee, Michelle y Singh, Jatinder. «Reviewable Automated Decision-Making: A Framework for Accountable Algorithmic Systems». *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, (2021), p. 598-609. DOI: https://doi.org/10.1145/3442188.3445921

Comité Europeo de Protección de Datos. «Coordinated Enforcement Action, Designation and Position of Data Protection Officers», (17 de enero de 2024) (en línea) [Fecha de consulta: 01.09.2024] https://www.edpb.europa.eu/our-work-tools/our-documents/other/coordinated-enforcement-action-designation-and-position-data_en

Crootof, Rebecca; Kaminski, Margot E. y Price II, William Nicholson. «Humans in the Loop». *Vanderbilt Law Review*, vol. 76, n.º 2 (2023), p. 429-510.

Custers, Bart y Heijne, Anne-Sophie. «The Right of Access in Automated Decision-Making: The Scope of Article 15(1)(h) GDPR in theory and practice». *Computer Law and Security Review*, (2022) DOI: https://doi.org/10.1016/j.clsr.2022.105727

Dreyer, Stephan y Schulz, Wolfgang. «The General Data Protection Regulation and Automated Decision-making: Will it deliver?». *Discussion Paper Ethics of Algorithms*, n.º 5, Bertelsmann Stiftung, (2019) (en línea) DOI: https://doi.org/10.11586/2018018

Edwards, Lilian y Veale, Michael. «Enslaving the Algorithm: From a 'Right to an Explanation' to a 'Right to Better Decisions'?». *IEEE Security & Privacy*, vol. 16, n.º 3 (2018), p. 46-54. DOI: https://doi.org/10.1109/MSP.2018.2701152

Enarsson, Therese; Enqvist, Lena y Naarttijärvi, Markus. «Approaching the human in the loop – legal perspectives on hybrid human/algorithmic decision-making in three contexts». *Information & Communications Technology Law*, vol. 31, n.º 1 (2021), p. 123-153. DOI: https://doi.org/10.1080/13600834.2021.1958860

Enqvist, Lena. «Human oversight in the EU artificial intelligence act: what, when and by whom?». *Law, Innovation & Technology*, vol. 15, n.º 2 (2023), p. 508-535. DOI: https://doi.org/10.1080/17579961.2023.2245683

Grant, David Gray; Behrends, Jeff y Basl, John. «What we owe to decision-subjects: beyond transparency and explanation in automated decision-making». *Philosophical Studies*, (2023). DOI: https://doi.org/10.1007/s11098-023-02013-6 (en línea) [Fecha de consulta: 01.09.2024] https://link.springer.com/article/10.1007/s11098-023-02013-6

Green, Ben. «The flaws of policies requiring human oversight of government algorithms». *Computer Law & Security Review*, vol. 45, (2022). DOI: https://doi.org/10.1016/j.clsr.2022.105681 (en línea) [Fecha de consulta: 01.09.2024] https://www.sciencedirect.com/science/article/pii/S0267364922000292?via%3Dihub

Kaminski, Margot E. y Malgieri, Gianclaudio. «Algorithmic impact assessments under the GDPR: Producing multi-layered explanations». *International Data Privacy Law*, vol. 11, n.º 2 (2021), p. 125-144. DOI: https://doi.org/10.1093/idpl/ipaa020

Kern, Christoph; Gerdon, Frederic; Bach, Ruben L.; Keusch, Florian y Kreuter, Frauke. «Humans versus machines: Who is perceived to decide fairer? Experimental evidence on attitudes toward automated decision-making». *Patterns*, vol. 3, n.º 10 (2022), p. 1-12. DOI: https://doi.org/10.1016/j.patter.2022.100591

Koivisto, Ida; Koulu, Riikka y Larsson, Stefan. «User accounts: How technological concepts permeate public law through the EU's AI Act». *Maastricht*

Journal of European and Comparative Law, vol. 0, n.º 0 (2024). DOI: https://doi.org/10.1177/1023263X241248469 [Fecha de consulta: 01.09.2024] https://journals.sagepub.com/doi/10.1177/1023263X241248469

Koulu, Riikka. «Proceduralizing control and discretion: Human oversight in artificial intelligence policy». *Maastricht Journal of European and Comparative Law*, vol. 27, n.º 6 (2020), p. 720-735. DOI: https://doi.org/10.1177/1023263X20978649

Kyriakou, Kyriakos y Otterbacher, Jahna. «In humans, we trust». *Discover Artificial Intelligence*, vol. 3, n.º 44 (2023), p. 1-18. DOI: https://doi.org/10.1007/s44163-023-00092-2

Lazcoz, Guillermo y De Hert, Paul. «Humans in the GDPR and AIA governance of automated and algorithmic systems. Essential pre-requisites against abdicating responsibilities». *VUB Brussels Privacy Hub Working Paper*, vol. 8, n.º 32 (2022), p. 1-28. DOI: https://doi.org/10.2139/ssrn.4016502

Logg, Jennifer M.; Minson, Julia A. y Moore Don A. «Algorithm appreciation: people prefer algorithmic to human judgment». *Organizational Behavior & Human Decision Processes*, vol. 151, (2019), p. 90-103. DOI: https://doi.org/10.1016/j.obhdp.2018.12.005

Lynskey, Orla. «Regulating for the Future: The Law's Enforcement Deficit». *Studies: An Irish Quarterly Review*, vol. 112, n.º 445 (2023), p. 104-119. DOI: https://doi.org/10.1353/stu.2023.0007

Mahieu, René L. P. y Ausloos, Jef. «Recognising and Enabling the Collective Dimension of the GDPR and the Right of Access». *Law Archive Papers*, (29 de abril de 2020a), p. 3-38. DOI: https://doi.org/10.31228/osf.io/b5dwm

Mahieu, René L. P. y Ausloos, Jef. «Harnessing the collective potential of GDPR access rights: towards an ecology of transparency». *Internet Policy Review*, (6 de julio de 2020b) (en línea) [Fecha de acceso: 01.09.2024] https://policyreview.info/articles/news/harnessing-collective-potential-gdpr-access-rights-towards-ecology-transparency/1487

Malgieri, Gianclaudio. «'Just' Algorithms: Justification (Beyond Explanation) of Automated Decisions Under the General Data Protection Regulation». *Law and Business*, vol. 1, n.º 1 (2021), p. 16-28. DOI: https://doi.org/10.2478/law-2021-0003

Mantelero, Alexander. *Beyond Data Human Rights, Ethical and Social Impact Assessment in AI*. La Haya: T.M.C Asser Press, 2022.

Mendoza, Isak y Bygrave, Lee A. «The Right not to be Subject to Automated Decisions based on Profiling», en: Synodinou, Tatiana-Eleni; Jougleux, Philippe; Markou, Christina y Prastitou, Thalia (eds.) *EU Internet Law: Regulation and Enforcement*. Cham: Springer, 2017, p. 77-98.

Misuraca, Gianluca y van Noordt, Colin. «AI Watch: Artificial Intelligence in public services, EUR 30255 EN». *Publications Office of the European Union*, (2020). DOI: https://doi.org/10.2760/039619, JRC120399

Novelli, Claudio; Taddeo, Mariarosaria y Floridi, Luciano. «Accountability in artificial intelligence: what it is and how it works». *AI & Society*, vol. 39, (2024), p. 1.871-1.882. DOI: https://doi.org/10.1007/s00146-023-01635-y

Parlamento Europeo. «Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance)». *Official Journal of the European Union*, L series, 2024/1689, (12 de julio de 2024) (en línea) http://data.europa.eu/eli/reg/2024/1689/oj

Roig, Antoni. «Safeguards for the right not to be subject to a decision based solely on automated processing (Article 22 GDPR)». *European Journal of Law and Technology*, vol. 8, n.º 3 (2017), p. 1-17 (en línea) https://ejlt.org/index.php/ejlt/article/view/570

Roig, Antoni. L*as garantías frente a las decisiones automatizadas. Del Reglamento General de Protección de Datos a la gobernanza algorítmica*. Barcelona: J.M. Bosch, 2020.

Rovatsos, Michael; Mittelstadt, Brent y Koene, Ansgar. «Landscape Summary: Bias in Algorithmic Decision-Making: What is bias in algorithmic decision-making, how can we identify it, and how can we mitigate it?». *UK Government*, Research and analysis, (19 de julio de 2019) (en línea) [Fecha de consulta: 01.09.2024] https://www.gov.uk/government/publications/landscape-summaries-commissioned-by-the-centre-for-data-ethics-and-innovation

Sartor, Giovanni y Lagioia, Francesca. «The impact of the General Data Protection Regulation (GDPR) on artificial intelligence». *Panel for the Future of Science and Technology (STOA)*, (25 de junio de 2020) (en línea) [Fecha de consulta: 01.09.2024] https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2020)641530

Schmidt, Philipp; Biessmann, Felix y Teubner, Timm. «Transparency and trust in artificial intelligence systems». *Journal of Decision Systems*, vol. 29, n.º 4 (2020), p. 260-278. DOI: https://doi.org/10.1080/12460125.2020.1819094

Selbst, Andrew D. y Powles, Julia. «Meaningful information and the right to explanation». *International Data Privacy Law*, vol. 7, n.º 4 (2017), p. 233-242. DOI: https://doi.org/10.1093/idpl/ipx022

Sivan-Sevilla, Ido. «Varieties of enforcement strategies post-GDPR: a fuzzy-set qualitative comparative analysis (fsQCA) across data protection authorities».

Journal of European Public Policy, vol. 31, n.º 2 (2024), p. 552-585. DOI: https://doi.org/10.1080/13501763.2022.2147578

Sterz, Sarah; Baum, Kevin; Biewer, Sebastian; Hermanns, Holger; Lauber-Rönsberg, Anne; Meinel, Philip y Langer, Markus. «On the quest for effectiveness in human oversight: Interdisciplinary perspectives». *FAccT '24: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, (2024), p. 2.495-2.507. DOI: https://doi.org/10.1145/3630106.3659051

Tsamados, Andreas; Floridi, Luciano y Taddeo, Mariarosaria. «Human control of AI systems: from supervision to teaming». *AI Ethics*, (2024). DOI: https://doi.org/10.1007/s43681-024-00489-4

Wagner, Ben. «Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision Making Systems». *Policy and Internet*, vol. 11, n.º 1 (2019), p. 104-122. DOI: https://doi.org/10.1002/poi3.198

Wieringa Maranke. «"Hey SyRI, tell me about algorithmic accountability": Lessons from a landmark case». *Data & Policy*, vol. 5, (2023), p. 1-24. DOI: https://doi.org/10.1017/dap.2022.39

WP251, European Data Protection Board. «Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679», 17/EN WP251rev.01, (3 de octubre de 2017) (en línea) [Fecha de acceso: 01.09.2024] https://ec.europa.eu/newsroom/article29/items/612053/en

Yeung, Karen. «Algorithmic Regulation: A Critical Interrogation». *Regulation and Governance*, vol. 12, n.º 4 (2018), p. 505-523. DOI: https://doi.org/10.1111/rego.12158

Traducción del original en inglés: Camino Villanueva, Massimo Paolini y redacción CIDOB.