

Estudios de Deusto

Revista de Derecho Público

Vol. 72/1 enero-junio 2024

DOI: <https://doi.org/10.18543/ed7212024>

ESTUDIOS

LA INTELIGENCIA ARTIFICIAL: DEFINICIÓN, REGULACIÓN Y RIESGOS PARA LOS DERECHOS FUNDAMENTALES

*Artificial Intelligence: Definition,
and Risks to Fundamental Rights*

María Pérez-Ugena

Profesora Titular de Derecho Constitucional.

Universidad rey Juan Carlos, España

<https://orcid.org/0000-0002-2724-6882>

<https://doi.org/10.18543/ed.3108>

Fecha de recepción: 20.12.2023

Fecha de aceptación: 10.05.2024

Fecha de publicación en línea: junio 2024

Derechos de autoría / Copyright

Estudios de Deusto. Revista de Derecho Público es una revista de acceso abierto, lo que significa que es de libre acceso en su integridad. Se permite su lectura, la búsqueda, descarga, distribución y reutilización legal en cualquier tipo de soporte sólo para fines no comerciales, sin la previa autorización del editor o el autor, siempre que la obra original sea debidamente citada y cualquier cambio en el original esté claramente indicado.

Estudios de Deusto. Revista de Derecho Público is an Open Access journal which means that it is free for full access, reading, search, download, distribution, and lawful reuse in any medium only for non-commercial purposes, without prior permission from the Publisher or the author; provided the original work is properly cited and any changes to the original are clearly indicated.

Estudios de Deusto

© Universidad de Deusto • ISSN 0423-4847 • ISSN-e 2386-9062, Vol. 72/1, enero-junio 2024

<http://www.revista-estudios.deusto.es/>

LA INTELIGENCIA ARTIFICIAL: DEFINICIÓN, REGULACIÓN Y RIESGOS PARA LOS DERECHOS FUNDAMENTALES

Artificial Intelligence: Definition, Regulation, and Risks to Fundamental Rights

María Pérez-Ugena

Profesora Titular de Derecho Constitucional.

Universidad rey Juan Carlos, España

<https://orcid.org/0000-0002-2724-6882>

<https://doi.org/10.18543/ed.3108>

Fecha de recepción: 20.12.2023

Fecha de aceptación: 10.05.2024

Fecha de publicación en línea: junio 2024

Resumen

Este trabajo se propone definir el alcance de la inteligencia artificial, a pesar de las dificultades y la falta de consenso que rodean este concepto. El objetivo principal es analizar los efectos de la regulación de la inteligencia artificial y sus consecuencias en los derechos fundamentales desde una perspectiva general. Se examina en detalle el estado actual de la regulación de la inteligencia artificial, con especial atención al Reglamento de la Unión Europea y al Acuerdo del 8 de diciembre de 2023.

Se ponen en relación la inteligencia artificial y los derechos fundamentales, mediante el análisis de aquellos derechos que podrían verse más claramente amenazados por los sistemas de inteligencia artificial con el objetivo final de lograr una visión global del impacto de la inteligencia artificial en el régimen de derechos. Este análisis busca, en definitiva, contribuir a la comprensión y discusión informada sobre los aspectos éticos y legales que la inteligencia artificial plantea en la actualidad.

Palabras clave

Inteligencia artificial; regulación; derechos fundamentales.

Abstract

This paper aims to define the scope of artificial intelligence, despite the challenges and lack of consensus surrounding this concept. The main objective is to analyze the effects of the regulation of artificial intelligence and its consequences on fundamental rights from a general perspective. The current state of artificial intelligence regulation is examined in detail, with special attention to the European Union's Regulatory and the Agreement of December 8, 2023.

The relationship between artificial intelligence and fundamental rights is explored by analyzing those rights that may be more clearly threatened by artificial intelligence systems. The goal is to achieve a comprehensive understanding of the impact of artificial intelligence on the rights framework. This analysis seeks, in essence, to contribute to an informed understanding and discussion of the ethical and legal aspects that artificial intelligence poses today.

Keywords

Artificial intelligence; regulation; fundamental rights.

Sumario: I. INTRODUCCIÓN II. DEFINICIÓN Y TIPOS DE SISTEMAS DE INTELIGENCIA ARTIFICIAL. 1. Definición de sistemas IA. 2. Tipologías de sistemas IA. 3. El algoritmo como pieza clave de sistemas IA. III. REGULACIÓN DE LA IA. 1. Antecedentes de la actual regulación. 2. El Reglamento de IA de la UE y el acuerdo de 8 de diciembre. IV. DECLARACIONES Y ACUERDOS INTERNACIONALES BASADOS EN LA ÉTICA. V. LOS DERECHOS FUNDAMENTALES FRENTE A LA INTELIGENCIA ARTIFICIAL. VI. CONCLUSIONES. VII. REFERENCIAS BIBLIOGRÁFICAS.

I. INTRODUCCIÓN¹

La inteligencia artificial (IA) es una tecnología que tiene el potencial de transformar la sociedad y el orden jurídico-constitucional, al afectar a diversos ámbitos y derechos fundamentales. El objetivo de este trabajo es definir los sistemas IA, así como analizar el marco normativo existente y las dificultades que se presentan para su armonización global. Para ello, se parte de una definición de los sistemas de IA basada en sus capacidades, que permite abarcar la diversidad y la evolución de los desarrollos tecnológicos. A continuación, se examina la regulación actual de la IA y se identifican los principales problemas y lagunas que se plantean. Asimismo, se incide en los riesgos que suponen los sistemas de IA para los derechos fundamentales. Se hace especial referencia a la igualdad, los derechos del ámbito de la privacidad, la libertad de expresión, la participación y la tutela judicial efectiva.

Con carácter general, la tecnología, salvo en supuestos muy concretos, no ha creado problemas nuevos, sino que ha incidido en los ya existentes y ha hecho que éstos adquieran ciertas nuevas dimensiones que hacen precisas nuevas políticas y decisiones. Sin embargo, la inteligencia artificial ha irrumpido en la sociedad produciendo cambios tan relevantes, que exigen respuestas y políticas globales y actualizadas. Se produce una verdadera transformación del sistema y del orden jurídico constitucional, que obliga a una reinterpretación de éste desde nuevos postulados.

La IA afecta al régimen de derechos fundamentales. Prácticamente en la totalidad de ellos resulta de interés estudiar el impacto que produce, en cada uno, el uso la inteligencia artificial, desde la igualdad, a los derechos del ámbito de la vida privada, especialmente lo relacionado con la protección de datos, hasta derechos del ámbito de la comunicación, la libertad de expresión y a los derechos relacionados con la tutela judicial efectiva.

¹ Trabajo realizado dentro del proyecto de investigación: La libertad de expresión del juez en el marco del respeto a los principios de ética judicial en España pid2021-127122nb-i00

Para realizar este estudio, se ha seguido un método de investigación jurídica, basado en el análisis de fuentes normativas, doctrinales y jurisprudenciales, tanto nacionales como internacionales, que abordan la cuestión de la IA desde diferentes perspectivas. Se ha prestado especial atención a los documentos y propuestas de la Unión Europea, que se ha posicionado como un actor clave en el desarrollo de una IA ética y responsable.

II. DEFINICIÓN Y TIPOS DE SISTEMAS DE INTELIGENCIA ARTIFICIAL

1. *Definición de sistemas IA*

La inteligencia artificial² (IA) es una tecnología que ha experimentado un avance espectacular en poco tiempo, gracias a la combinación de factores como el *big data*, el *blockchain*, la nube, el internet de las cosas, la robótica y la realidad virtual. Aunque la IA no es una invención reciente, ya que sus orígenes se remontan a hace más de 50 años, su impacto actual es enorme y afecta a casi todos los ámbitos de la vida. Una conjunción de factores, como los avances en la potencia informática, la disponibilidad de enormes cantidades de datos y nuevos algoritmos han permitido que se produzcan grandes logros en los sistemas IA en los últimos años³.

No hay una definición clara de estos sistemas que goce de amplio consenso, porque, de una parte, la IA está sometida a las variaciones que se produzcan como consecuencia de los avances tecnológicos, de forma que no responde a algo estático, sino que es producto de una tecnología disruptiva, que además está desarrollándose a pasos acelerados. De otra, no hay un consenso en los agentes implicados sobre lo que son los sistemas IA. Sin embargo, con el objeto de regular la inteligencia artificial (IA) de manera efectiva es necesaria una comprensión común de lo que se entiende por ‘inteligencia artificial’. Con ello se pretende dar respuesta al interrogante de qué es lo que se quiere regular y porqué. Cuáles son los aspectos que se consideran “peligrosos” y que deban ser objeto de regulación. No todos los sistemas IA son objeto de preocupación porque no todos pueden afectar al régimen de derechos.

² El término IA, Inteligencia Artificial, fue usado por primera vez en 1956 por John McCarthy para referirse a “*la ciencia y la ingeniería de crear máquinas inteligentes especialmente programas de computación inteligentes*”. La RAE lo define como “*Disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico*”.

³ Parlamento Europeo. ¿Qué es la inteligencia artificial y cómo se usa?

<https://www.europarl.europa.eu/news/es/headlines/priorities/inteligencia-artificial-en-la-ue/20200827STO85804/que-es-la-inteligencia-artificial-y-como-se-usa>

Se trata de una cuestión compleja, pues no es un solo fenómeno, sino que hay muchas formas de inteligencia artificial. Esto obliga a buscar aproximarse al concepto de inteligencia artificial a través de sus características predominantes y diferenciando entre los distintos tipos existentes. En este sentido, lo que define a los sistemas de IA es la capacidad de procesar datos e información de una manera que se asemeja a un comportamiento inteligente, y abarca generalmente aspectos de razonamiento, aprendizaje, creatividad o la capacidad de predecir o planear. Es de un conjunto de técnicas, por tanto, que permiten a un sistema informático simular características que son propias de la inteligencia humana. De manera que se automatizan ciertas actividades hasta ahora vinculadas exclusivamente con procesos del pensamiento humano, como pueden ser la toma de decisiones.

Los sistemas de IA son capaces de adaptar su comportamiento, en cierta medida, analizar los efectos de acciones previas y trabajar de manera autónoma. Son tecnologías de procesamiento de la información que integran modelos y algoritmos con capacidad para aprender y realizar tareas cognitivas, dando lugar a resultados como la predicción y la adopción de decisiones en entornos materiales y virtuales. La inteligencia artificial se relaciona de forma clara con el *big data*. Lo necesita para desarrollar sus funcionalidades, ya que se nutre de la gran cantidad de datos recopilados para entrenar modelos de aprendizaje automático y tomar decisiones basadas en patrones y correlaciones. Esta sinergia permite a la inteligencia artificial realizar tareas como el procesamiento de lenguaje natural, la visión por computadora y la toma de decisiones predictivas con un alto grado de precisión (Cotino Hueso, L. 2017). La tecnología *blockchain* también desempeña un papel importante en este ecosistema (Merchán Murillo, A. 2019). Al aprovechar la seguridad y la inmutabilidad que proporciona la cadena de bloques, tanto el *big data* como la inteligencia artificial pueden garantizar la integridad de los datos, lo que es esencial en aplicaciones críticas como la autenticación de identidades digitales y en la gestión de claves criptográficas, lo que fortalece la seguridad en las comunicaciones y transacciones digitales. La ciberseguridad es otra cuestión relevante en este ámbito⁴.

⁴ El Convenio de Budapest, oficialmente conocido como el Convenio sobre Delitos Cibernéticos, representa un hito fundamental en la lucha contra las infracciones informáticas. Este tratado pionero fue concebido por el Consejo de Europa en Estrasburgo, con la colaboración activa de observadores de estados como Canadá, Japón y China, marcando así un esfuerzo conjunto para abordar las amenazas cibernéticas.

Aprobado por el Comité de ministros del Consejo de Europa en su 109.^a reunión el 8 de noviembre de 2001, el Convenio de Budapest se abrió a la firma en Budapest el 23 de noviembre de 2001, y finalmente entró en vigencia el 1 de julio de 2004; El 1 de marzo de 2006, entró en vigor el Protocolo Adicional a la Convención sobre Delitos Cibernéticos; El segundo Protocolo Adicional de la Convención de Budapest, firmado el 12 de mayo de

La definición que realiza el Proyecto de Reglamento sobre IA del Parlamento Europeo y del Consejo de 2021 se sustenta en sus capacidades. En qué es lo que pueden hacer los sistemas IA. Así, señala que *los sistemas de IA vendrían definidos por el software que se desarrolla, empleando ciertas técnicas o estrategias, y que puede, para un conjunto determinado de objetivos definidos por seres humanos, generar información de salida como contenidos, predicciones, recomendaciones o decisiones que influyan en los entornos con los que interactúa*⁵.

A través de estas técnicas o estrategias, la inteligencia artificial permite que los sistemas tecnológicos perciban su entorno, se relacionen con él, resuelvan problemas y actúen con un fin específico. De forma que, el sistema recibe datos, bien sean ya preparados o recopilados a través de sus propios sensores, los procesa y responde a ellos.

Otra aproximación en el intento de definir los sistemas IA se fundamenta en dos de sus características, que son las que generan la necesidad de una respuesta regulatoria específica⁶. Se refiere a la explicabilidad de la IA, de una parte, y a su autonomía, de otra. La explicabilidad es la capacidad de entender y comunicar cómo y por qué los sistemas de IA producen ciertos resultados. La IA se basa en el aprendizaje a partir de los datos, y este aprendizaje puede conducir a formas de inferencia que no son directamente programadas por los humanos, sino que son descubiertas por los propios sistemas de IA.

Esto puede hacer difícil explicar la intención o la lógica de los resultados del sistema, especialmente cuando se trata de modelos complejos como los de *deep learning*, que utilizan redes neuronales artificiales con muchas capas y parámetros (Arellano Toledo 2019, W.; Cotino Hueso, L y Castellanos Claramunt, J 2022).

La explicabilidad deriva en una necesaria regulación de la IA⁷ que garantice a las personas poder conocer los motivos por los que se ha

2022, se centra en proporcionar herramientas efectivas para mejorar la cooperación y la divulgación de evidencia electrónica,

⁵ Propuesta de Reglamento del Parlamento europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de inteligencia artificial) y se modifican determinados actos legislativos de la Unión, Bruselas, 21 de abril de 2021, COM (2021) 206 final. En un sentido parecido, la Recomendación sobre la ética de la Inteligencia artificial <https://www.unesco.org/es/legal-affairs/recommendation-ethics-artificial-intelligence> UNESCO. 2021. “Técnica de procesamiento de la información integra modelos y algoritmos con capacidad para aprender tareas cognitivas y ofrece como resultados la predicción y la adopción de decisiones”.

⁶ *Policy paper A pro-innovation approach to AI regulation Updated 3 August 2023* <https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper>

⁷ *¿Qué es la IA explicable?* <https://www.ibm.com/es-es/topics/explainable-ai>

tomado una decisión que afecta a sus derechos y libertades y tener la posibilidad de recurrir las decisiones (Samek, W. Montavon, G. Vedaldi, A., Kai Hansen L, Müller K.R. 2019). Los actores de la IA deberían informar a los usuarios cuando un producto o servicio se proporcione directamente o con la ayuda de sistemas de IA de manera adecuada y oportuna, en particular cuando afecta a su seguridad o a sus derechos humanos; en esas circunstancias, se debería tener la oportunidad de solicitar explicaciones e información al actor de la IA o a las instituciones del sector público correspondientes sobre los factores que influyen en una predicción o decisión específicas, y sobre la existencia o no de garantías adecuadas (como medidas de seguridad o de equidad)⁸.

La explicabilidad está estrechamente relacionada con la transparencia, ya que los resultados y los subprocesos que conducen a ellos deberían aspirar a ser comprensibles y trazables, apropiados al contexto. Los actores de la IA deberían comprometerse a velar por que los algoritmos desarrollados sean explicables. En el caso de las aplicaciones de IA cuyo impacto en el usuario final no es temporal, fácilmente reversible o de bajo riesgo, debería garantizarse que se proporcione una explicación satisfactoria con toda decisión que haya dado lugar a la acción tomada, a fin de que el resultado se considere transparente⁹.

La autonomía de la inteligencia artificial se refiere al grado de independencia y autocontrol que poseen los algoritmos al llevar a cabo una tarea sin necesidad de intervención o supervisión humana. Desde la ejecución de instrucciones pre-programadas hasta la capacidad de generar acciones propias basadas en el aprendizaje y la adaptación al entorno, la autonomía de la IA abarca un espectro amplio. (Arellano Toledo 2019, W.; Cotino Hueso, L y Castellanos Claramunt, J 2022).

Estos elementos no solo pueden tener un carácter impredecible, sino que también poseen la capacidad de infligir daño físico, marcando así el inicio de una nueva fase en la interacción entre los seres humanos y la tecnología. A diferencia del software y de Internet, la particularidad de la IA radica en su capacidad para interactuar con el mundo fuera de línea, manifestando su influencia de manera física en el entorno. Esta autonomía plantea cuestiones importantes, ya que la capacidad de la IA para actuar físicamente en el mundo real implica un potencial real de causar daño material a personas u objetos (Moisés Barrio A. 2018). Así como la dificultad para asignar responsabilidad por los resultados, ya que la autonomía de la IA puede llevar a que

⁸ Recomendación sobre la ética de la inteligencia artificial, adoptada el 23 de noviembre de 2021 por la UNESCO

⁹ *Ibid*

algunos sistemas tomen decisiones sin la intención expresa o el control continuo de un ser humano¹⁰.

Para abordar estas cuestiones, se ha propuesto enfoques basados en las capacidades funcionales de la IA, es decir, en lo que puede hacer la IA y cómo lo hace, en lugar de en las tecnologías o aplicaciones específicas de la IA, como el reconocimiento facial o los modelos de lenguaje. De esta manera, se trata de hacer valer un principio de neutralidad tecnológica y se evita que el marco regulatorio se quede obsoleto o limitado por las definiciones legales rígidas, que pueden no abarcar las nuevas formas de IA que puedan surgir en el futuro¹¹

2. Tipologías de sistemas IA

Respecto de las tipologías de sistemas IA, no solo existen distintos tipos, sino que, además, no todos evolucionan de la misma forma. De hecho, se ha producido un gran avance en lo que se denomina IA estrecha, encargada de hacer una tarea concreta, pero no en la IA general, en la que los logros han sido menores comparativamente¹². La IA estrecha, también conocida como IA débil, se refiere a sistemas diseñados para realizar tareas específicas y limitadas, como el reconocimiento de voz, la identificación de imágenes o la traducción de idiomas. No tienen capacidad de aprendizaje o adaptación por sí mismos, y requieren ser programados para realizar una tarea determinada. Su alcance es limitado y no pueden realizar tareas fuera de su campo de especialización. La IA fuerte, en cambio, está diseñada para tener una amplia gama de habilidades cognitivas y capacidad de aprendizaje autónomo. Estos sistemas pueden realizar múltiples tareas y aprenden de forma autónoma a medida que interactúan con el entorno.

La IA fuerte se caracteriza por tener capacidad de razonar, planificar y tomar decisiones complejas en un amplio espectro de situaciones; La IA superinteligente sería otro tipo de IA que superaría la inteligencia humana en todos los aspectos. Este nivel de IA sería capaz de comprender el mundo de una manera que está más allá de la capacidad humana, y sería capaz de

¹⁰ Vid informe del Parlamento Europeo con recomendaciones destinadas a la Comisión Europea sobre un régimen de responsabilidad civil en materia de IA, aprobado en octubre de 2020 https://www.europarl.europa.eu/doceo/document/A-9-2020-0178_EN.pdf

¹¹ *Policy paper A pro-innovation approach to AI regulation Updated 3 August 2023* <https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper>

¹² *¿Qué es la Inteligencia Artificial? Plan de recuperación, transformación y resiliencia*, 2023. <https://planderecuperacion.gob.es/noticias/que-es-inteligencia-artificial-ia-prtr>

resolver problemas complejos a una velocidad y eficiencia que los seres humanos no pueden alcanzar. Es una forma teórica de IA que aún no ha sido desarrollada en la práctica.

La inteligencia artificial no solo ofrece oportunidades innovadoras, sino que también plantea riesgos significativos. La capacidad de procesamiento masivo de datos y los algoritmos de aprendizaje automático de la IA pueden generar decisiones automatizadas con un impacto directo en los individuos y en la sociedad en su conjunto. Por lo tanto, se hace necesario garantizar que estas decisiones no vulneren derechos fundamentales ni afecten a los principios constitucionales básicos.

Los sistemas de IA pueden utilizarse en los sistemas ciberfísicos, incluidos la Internet de las cosas, que se refiere a la tecnología basada en la conexión de objetos a la red, desde electrodomésticos a coches autónomos y ciudades inteligentes, los sistemas robóticos, la robótica social y las interfaces entre seres humanos y ordenadores, que comportan el control, la percepción, el procesamiento de los datos recogidos por sensores y el funcionamiento de los actuadores en el entorno en que operan los sistemas de IA.

3. *El algoritmo como pieza clave de sistemas IA*

La pieza clave en los sistemas IA es el algoritmo. Elemento fundamental de todo el engranaje que, al expresarse en un lenguaje de programación, permite que se realicen actividades registradas y procesadas por sistemas inteligentes. Las Inteligencias artificiales utilizan algoritmos para procesar grandes cantidades de datos y tomar decisiones basadas en patrones y reglas establecidas a través del aprendizaje automático.

Puede definirse como un “conjunto de instrucciones o reglas definidas y no-ambiguas, ordenadas y finitas que permite, típicamente, contestar una pregunta, tomar una decisión, solucionar un problema, realizar un cómputo, procesar datos o llevar a cabo alguna tarea”. Estos procedimientos computacionales toman uno o varios valores de entrada y generan uno o varios valores de salida, por lo tanto, son instrumentos que no intentan establecer un vínculo causal entre una variable específica y su efecto, sino que producen un resultado”¹³.

Se refiere a tipos de sistemas diferentes, de acuerdo con variables como los datos que el mismo maneja, del tipo de funcionamiento interno que se establece para dicho sistema, o de los objetivos de su funcionamiento, entre otros. Desde la perspectiva de su ámbito de acción, hay infinidad de tipos de algoritmos, los que interesan desde la óptica del Derecho Público son

¹³ *Guía de auditoría algorítmica. Éticas-consulting*. <https://www.eticasconsulting.com/eticas-consulting-guia-de-auditoria-algoritmica-para-desarrollar-algoritmos-justos-y-eficaces/>

aquellos que afectan a sectores públicos o de interés social; aquellos que guardan relación con el acceso a ciertas prestaciones que se relacionan con los derechos fundamentales o con bienes constitucionalmente protegidos. Es decir, cuestiones como el acceso a la educación o al trabajo, otras que conlleven discriminación basada en algún posible aspecto derivado del artículo 14 de la Constitución, o que se implementan en el ámbito judicial.

El algoritmo va a depender de cómo lo hayan previsto sus diseñadores, desarrolladores y operadores. En el trabajo de buscadores se accede a la información según ésta se haya categorizado. De forma que los resultados se muestran no solo según las peticiones de los usuarios sino según se hayan categorizado. Básicamente, el problema que se plantea con el uso de los algoritmos, especialmente aquellos de aprendizaje automático, a los que se ha definido como “cajas negras” es la dificultad de que las decisiones derivadas puedan ser comprendidas o controladas, al convertirse en conjuntos opacos de códigos informáticos y datos. Por eso, resulta fundamental que los algoritmos se diseñen, desarrollen y utilicen de forma que sean, no solo adecuados desde el punto de vista jurídico, sino también deben ser fáciles de supervisar y corregir, controlando cualquier sesgo. La transparencia se convierte en un aspecto fundamental para prevenir, identificar y mitigar posibles sesgos¹⁴.

Se puede precisar que el objetivo de análisis a la hora de estudiar cómo se afecta a los derechos fundamentales, no son los algoritmos en sí mismos, sino los procesos de toma de decisiones en torno a los algoritmos.

Se plantea, por último, cual es la naturaleza jurídica de los algoritmos y si éstos pueden ser considerados como una fuente del derecho. A estos efectos, es preciso tener en cuenta que los acuerdos o contratos, a diferencia de las normas jurídicas, carecen de efectos *erga omnes*, es decir, no se aplican universalmente, sino que sus consecuencias se limitan a las partes que los suscriben. Esto va en contra de la característica fundamental de las leyes, que es su aplicabilidad generalizada. Incluso cuando un contrato emitido por una compañía tecnológica afecta a millones de personas, la cantidad de suscriptores no modifica la esencia del algoritmo, el cual sigue confinado, en este caso, a un ámbito privado. En consecuencia, no puede ser considerado una norma jurídica ni formar parte del sistema de fuentes legales (Balaguer Callejón F. 2023)

También interesan algunas cuestiones en el desarrollo del derecho digital relativas a cómo se integran los algoritmos en el panorama legal. O si se requiere desarrollar nuevas categorías legales para abordar los efectos de los algoritmos en la sociedad (Balaguer Callejón F. 2023)

Respecto de la naturaleza jurídica del algoritmo, lo determinante debería ser el ámbito específico en el que se utiliza el algoritmo. De tal forma, que su

¹⁴ *Ibid.*

régimen jurídico no se haga depender de la naturaleza de los procesos técnicos que realizan ni del código fuente en sí mismo considerado. En el ámbito concreto de la Administración Pública (Gutiérrez David. ME, 2021, Vestri G. 2021) tampoco existe unanimidad acerca del carácter o naturaleza jurídica de los algoritmos, aunque puede aceptarse el planteamiento de que, si materialmente realizan las mismas funciones que los reglamentos, deben estar sometidos a las similares garantías (Boix Palop, A. 2020; Martín Delgado I. 2009). En lo que sí hay consenso es en la necesidad de que se regule, puesto que el marco jurídico existente en la actualidad no es suficiente ni permite que las decisiones tomadas sobre algoritmos sean trazables o explicables.

III. REGULACIÓN DE LA IA

El intento de crear un marco regulador de la IA es muy complejo, especialmente por el hecho de que los avances tecnológicos no son claramente previsibles y es preciso esperar a que se produzcan para que se plantee la necesidad de implementar normas jurídicas. Es decir, no hay una suficiente capacidad de anticipación en este sentido. De ahí que resulte imprescindible aplicar el Derecho que venga a proteger a la ciudadanía de una forma lo más abierta posible, en el sentido de lo más neutral. En última instancia, se busca que la regulación sea lo más resistente posible frente al avance del tiempo, considerando la rápida evolución tecnológica relacionada con la inteligencia artificial. En definitiva, se trata de enfrentarse a los problemas o riesgos que presentan la utilización de un desarrollo tecnológico atendiendo, principalmente al contenido, al fondo, al derecho que queda afectado, frente a la plataforma o técnica que se utilice, aun teniendo en cuenta que cada avance tecnológico crea nuevas posibilidades no previstas en la legislación con las consecuentes lagunas que se pueden producir.

Además, y puesto que la IA es sector hiper-regulado, se trata de buscar marcos generales reguladores, que aseguren un desarrollo y un uso de la IA ético, responsable y sostenible, puesto que la maraña normas aplicables, que actualmente regulan algunos usos de la inteligencia artificial (IA), puede que no resulte suficiente para enfrentar los riesgos que la IA puede plantear. (Hernández Peña, J.J. 2022)

1. *Antecedentes de la actual regulación*

Como antecedentes de su regulación, ética y gobernabilidad en el marco de la Unión Europea, además de la Carta de Derechos Fundamentales de la Unión Europea, a efectos de marco general regulador de derechos y principios inspiradores, como la dignidad, la libertad, la igualdad, la solidaridad, la

ciudadanía y la justicia, que son relevantes para el desarrollo y el uso de la IA, son referencias básicas las siguientes:

La Declaración de Tallin sobre Administración Electrónica. La Declaración de Tallin sobre Administración Electrónica de 2017 marcó un hito para crear una administración electrónica orientada a los servicios, fiable e innovadora en Europa, que pueda garantizar servicios públicos digitales transfronterizos de alta calidad, para los ciudadanos y las empresas en el desarrollo de un Mercado Único Europeo dirigido al futuro. El Plan de Acción sobre Administración Electrónica 2016-2020 va a establecer las prioridades y las acciones para modernizar la administración pública y facilitar la interacción con los ciudadanos y las empresas en la UE. El Plan incluye medidas para impulsar el uso de la IA y otras tecnologías digitales en la prestación de servicios públicos, así como para fomentar la colaboración y el intercambio de buenas prácticas entre los Estados miembros¹⁵

La Directiva de Accesibilidad Web establece los requisitos de accesibilidad que deben cumplir los sitios web y las aplicaciones móviles del sector público en la UE, para garantizar que sean accesibles para todas las personas, especialmente para las personas con discapacidad. La Directiva también promueve el uso de estándares abiertos y el desarrollo de soluciones innovadoras, como la IA, para mejorar la accesibilidad web¹⁶.

La también reciente Ley de gobernanza de datos está basada en la contribución voluntaria de datos para el bien común y en un conjunto de principios y prácticas que tienen como objetivo facilitar el acceso y el uso de los datos para fines de interés público, respetando los derechos y las preferencias de los proveedores de datos. Estas normas se fundamentan en la idea de que los datos son un recurso común que puede generar valor social, económico y ambiental, y que requiere una gestión responsable y participativa¹⁷.

La Declaración de Berlín sobre la sociedad digital y la administración digital basada en valores de 2020¹⁸, expresa el compromiso de los Estados

¹⁵ Comunicación de la Comisión Europea [COM(2016) 179 final]: Plan de Acción sobre Administración Electrónica de la UE <https://eur-lex.europa.eu/ES/legal-content/summary/accelerating-the-digital-transformation-of-governments-in-the-eu-2016-2020-action-plan.html>

¹⁶ Directiva (UE) 2016/2102 del Parlamento Europeo y del Consejo de 26 de octubre de 2016 sobre la accesibilidad de los sitios web y aplicaciones para dispositivos móviles de los organismos del sector público <https://www.boe.es/doue/2016/327/L00001-00015.pdf>

¹⁷ Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act) (Text with EEA relevance) <https://digital-strategy.ec.europa.eu/en/policies/data-governance-act>

¹⁸ Se adoptó en la reunión ministerial durante la presidencia alemana del Consejo de la UE el 8 de diciembre de 2020

miembros de la UE de avanzar hacia una sociedad digital y una administración digital basada en valores, que respete los derechos fundamentales, la democracia, el Estado de Derecho y la soberanía digital. La Declaración también destaca el papel de la IA como una herramienta para mejorar la calidad de vida y el bienestar de las personas, así como para afrontar los retos globales, como la pandemia de COVID-19 o el cambio climático¹⁹.

A nivel global, La UE ha desarrollado una estrategia digital, que se fundamenta en una serie de principios²⁰ y pretende una transformación digital protegida, segura y sostenible que sitúe a las personas en el centro, en consonancia con los valores y los derechos fundamentales de la UE. Con ese propósito, en febrero de 2020, la Comisión publicó el Libro Blanco sobre la inteligencia artificial: un enfoque europeo orientado a la excelencia y la confianza²¹. En el Libro Blanco se definen las opciones existentes para alcanzar el doble objetivo de promover la adopción de la IA y de abordar los riesgos vinculados a determinados usos de esta nueva tecnología

Recientemente, la Declaración Europea sobre los Derechos y Principios Digitales para la Década Digital expresa el compromiso de la Unión Europea con una transformación digital centrada en las personas, en consonancia con los valores fundamentales de la UE y los derechos fundamentales. Fue firmada por los presidentes del Parlamento Europeo, el Consejo y la Comisión el 15 de diciembre de 2022²². La Declaración contiene seis puntos claves, que abordan aspectos como la soberanía digital abierta, el respeto de los derechos fundamentales, la inclusión, la accesibilidad, la igualdad y la no discriminación, la participación en el espacio público digital, la seguridad, la protección y el empoderamiento, y la sostenibilidad en el entorno digital. La Declaración pretende ser un marco de referencia para el desarrollo de una IA ética y responsable, que garantice la protección de los derechos humanos y el bienestar de la sociedad.

¹⁹ High-Level Conference on a values-based digital transformation (incl. public panel discussion) <https://www.eu2020.de/eu2020-en/events/-/2357380>

²⁰ La Declaración, presentada por la Comisión en enero de 2022 señala una serie de principios relacionados con la transformación digital. https://wayback.archive-it.org/12090/*/https://ec.europa.eu/digital-single-market/en/news/digital-single-market-strategy-europe-com2015-192-final

²¹ Comisión Europea, *Libro Blanco sobre la inteligencia artificial: un enfoque europeo orientado a la excelencia y la confianza*, COM (2020) 65 final, 2020. <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52020DC0065>

²² Declaración Europea sobre los Derechos y Principios Digitales para la Década Digital (2023/C 23/01) [https://eurlex.europa.eu/legalcontent/ES/TXT/PDF/?uri=CELEX:32023C0123\(01\)&qid=1684248405089](https://eurlex.europa.eu/legalcontent/ES/TXT/PDF/?uri=CELEX:32023C0123(01)&qid=1684248405089)

2. El Reglamento de IA de la UE y el acuerdo de 8 de diciembre

En el marco de esa política digital, en abril de 2021, la Comisión Europea presentó una propuesta para establecer un marco normativo de la Unión Europea sobre inteligencia artificial (IA). Esta iniciativa, conocida como el proyecto de Ley de IA, tiene como objetivo principal regular de manera horizontal el campo de la IA. Resulta de un proceso de consulta iniciado en el año 2018, que ha venido trabajando con diversos actores para establecer un marco normativo. El régimen legal propuesto se centra en la utilización específica de los sistemas de IA y los riesgos asociados con el objetivo de lograr una actuación coordinada respecto de las implicaciones éticas y humanas de la IA, desde un enfoque humanista y centrado en la persona, especialmente en lo que respecta a la protección de los derechos fundamentales. Este enfoque es fundamental frente a los riesgos de que se produzcan arbitrariedades, sesgos, discriminaciones y otras vulneraciones de los Derechos fundamentales. (Cotino L. 2021)

La posición del regulador europeo se basa en los riesgos que supone el uso de la IA, de forma que tiene en cuenta la probabilidad de daños derivados, especialmente, de la inteligencia artificial generativa (Barrio Andrés, M. 2022). La UE pretende, con esta regulación, situarse como un líder mundial en el desarrollo de sistemas IA seguros y éticos²³. En diciembre de 2021, el Consejo de la UE acordó la posición general de los Estados miembros sobre la propuesta. Posteriormente, el 14 de junio de 2023, el Parlamento Europeo emitió su posición al respecto. El 8 de diciembre del mismo año se firmó un acuerdo entre el Parlamento Europeo y el Consejo sobre la, mal llamada, Ley de Inteligencia Artificial, que ha marcado un hito significativo en la regulación de esta tecnología a nivel mundial. Este acuerdo conlleva posibles implicaciones globales. Es probable que otras regiones consideren estas normativas como referencia al desarrollar sus propias leyes. Lo que podría contribuir a la formación de estándares globales para el desarrollo y la implementación de la inteligencia artificial²⁴.

En general, el Acuerdo representa un paso significativo hacia la creación de un marco regulatorio sólido para la inteligencia artificial y refleja las preocupaciones y dificultades asociados con su rápida evolución. Se estructura

²³ Comisión Europea (2021a): “Generar confianza mediante el primer marco jurídico sobre la IA”. En *Excelencia y confianza en la inteligencia artificial*. https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/excellence-trust-artificial-intelligence_es#generar-confianza-mediante-el-primer-marco-juridico-sobre-la-ia. Comisión Europea (2021b): Nuevas normas sobre la inteligencia artificial: preguntas y respuestas. https://ec.europa.eu/commission/presscorner/detail/es/QANDA_21_1683

²⁴ Sobre su contenido vid. <https://www.consilium.europa.eu/es/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/>

en cuatro categorías de riesgo: inaceptable, elevado, limitado y mínimo. Así, se acordó prohibir: a) “sistemas de categorización biométrica que utilicen características sensibles (por ejemplo, creencias políticas, orientación sexual o raza); b) la “extracción no dirigida de imágenes faciales” de Internet o grabaciones de circuito cerrado de televisión (CCTV) para crear bases de datos de reconocimiento facial; c) “sistemas de reconocimiento de emociones” en el lugar de trabajo y en instituciones educativas; d) “sistemas de crédito social” basados en el comportamiento social o características personales; y e) “sistemas que manipulen el comportamiento humano” para quebrantar su libre albedrío o que exploten las vulnerabilidades de las personas (por su edad, discapacidad o situación socio-económica)²⁵.

Junto con los sistemas prohibidos, se relacionan otros a los que se considera de riesgo mínimo y los de alto riesgo. Los primeros únicamente quedan sometidos a obligaciones básicas de transparencia, en particular la necesidad de garantizar la accesibilidad a su documentación técnica. Con carácter voluntario, eso sí, las empresas podrán someterse a códigos de conducta adicionales para estos sistemas de IA²⁶.

Los sistemas de IA clasificados como de alto riesgo forman parte de este grupo porque conllevan un posible daño a la salud, la seguridad, los derechos fundamentales, el medio ambiente, la democracia y el Estado de derecho. Se incluyen, también, los que se usen para influir en resultados electorales y la conducta de los votantes) que quedan sujetos, adicionalmente, a una serie de medidas. En primer lugar, la evaluación y mitigación de riesgos, de entre la cuales, exige una evaluación de impacto sobre derechos fundamentales. Además, exige las garantías de alta calidad de los datos empleados; Una medida que permita el control son los registros de actividad. Exige, igualmente medidas apropiadas de supervisión humana. Y, por último, la información y control ciudadano, en forma de explicaciones sobre las decisiones basadas en este tipo de sistemas que afecten a sus derechos y derecho a presentar quejas sobre este tipo de sistemas²⁷.

Respecto a la IA generativa, estos sistemas y los modelos sobre los que se basan (por ejemplo, los GPT de OpenAI) son asimismo objeto de regulación.

²⁵ De forma excepcional, se permite, en cambio, solo en determinados supuestos, el uso por las fuerzas de seguridad de sistemas biométricos en lugares públicos, siempre que se cuente con autorización judicial. Los supuestos, no obstante, que permiten el uso de este tipo de tecnología son los que tienen por finalidad localizar a la víctima de un delito muy grave, prevenir delitos de terrorismo o perseguir crímenes también muy graves. Este es uno de los temas más complicados, que ha exigido necesariamente la toma de la decisión mediante consenso.

²⁶ *Ibidem*.

²⁷ *Ibidem*.

Todos ellos deberán cumplir requisitos de transparencia que incluyen la elaboración de documentación técnica y la difusión de resúmenes detallados sobre el contenido digital utilizado en su entrenamiento, el cual, a su vez, habrá de respetar la normativa europea sobre derechos de autor²⁸.

Finalmente, para aquellos clasificados como “de alto impacto” por su potencial riesgo sistémico (en la misma línea de los citados para los sistemas de alto riesgo), y siempre que reúnan ciertos criterios, se disponen obligaciones más estrictas. Así, además de la obligación común con los sistemas de riesgo alto, de evaluar y mitigar tales riesgos sistémicos, deberán realizar pruebas contradictorias (el conocido como red teaming). Están obligados a informar a la Comisión sobre incidentes graves, así como sobre su eficiencia energética, y deben garantizar la ciberseguridad²⁹.

El acuerdo explica con claridad el estado actual el proceso legislativo y los plazos previstos para la entrada en vigor de la “Ley”, que deberá someterse a aprobación final por el Parlamento Europeo y por el Consejo. La totalidad del futuro Reglamento no será aplicable hasta dos años después, si bien algunas de sus disposiciones, como las relativas a las prohibiciones y a la IA generativa, serán aplicables al cabo de 6 y 12 meses desde la entrada en vigor, respectivamente. Se da la Comisión Europea un papel clave para la implementación de la Ley, ya que le permite adoptar actos delegados y de ejecución para especificar los requisitos técnicos y metodológicos de los sistemas de IA, así como para actualizar la lista de usos prohibidos o de alto riesgo. Esto implica que la Comisión tendrá que estar al día de los avances científicos y tecnológicos en el campo de la IA y consultar a las partes interesadas, como los proveedores, los usuarios, las autoridades nacionales y los expertos independientes³⁰.

El acuerdo también implica que los Estados miembros tendrán que adaptar sus legislaciones nacionales para garantizar la aplicación efectiva y coherente de la Ley, especialmente en lo que se refiere a las infracciones y sanciones. El acuerdo establece un marco común de sanciones administrativas, que pueden llegar hasta el 7% del volumen de negocios anual total de una empresa, pero deja a los Estados miembros la facultad de determinar las sanciones penales o civiles que consideren apropiadas. Además, los Estados miembros tendrán que designar a las autoridades nacionales competentes para supervisar y hacer cumplir la Ley, así como a los organismos notificados que se encargarán de evaluar la conformidad de los sistemas de IA de alto riesgo³¹.

²⁸ *Ibidem*.

²⁹ *Ibidem*.

³⁰ *Ibidem*.

³¹ Vid la nota “Alerta: Ley europea de Inteligencia Artificial.” <https://hsfnotes.com/madrid/2023/12/13/alerta-ley-europea-de-inteligencia-artificial/>

Pese a que este Reglamento se viene calificando como la primera Ley del mundo, en puridad no es así. Tanto China como los EE.UU han aprobado en 2023, normas generales al respecto. La norma europea, eso sí, lo es con arreglo a valores occidentales, en tanto que, en contraste con el Decreto presidencial estadounidense, es una norma a la vez horizontal e imperativa para todos sus destinatarios. Las diferencias regulatorias son claras. Lo más relevante es la tendencia a regular de acuerdo con defensa de la ética y la protección de los derechos fundamentales de la UE, de una forma más claramente intervencionista en Europa, frente al sistema de EEUU, que anunció su primera regulación de la inteligencia artificial en octubre de 2023, mediante una orden ejecutiva del presidente americano. La Orden establece una serie de principios y prácticas para el uso responsable y confiable de la IA por parte de las agencias federales, como la transparencia, la equidad, la rendición de cuentas, la seguridad y la privacidad. La Orden también crea un comité asesor de expertos del sector, la academia y la sociedad civil para supervisar la implementación de la regulación y proponer recomendaciones para mejorarla. La orden no se aplica a las empresas privadas ni a los usos militares de la IA. (Hernández Peña, J.J. 2022)³².

China publicó nuevas normas sobre la inteligencia artificial generativa (IA) en julio de 2023, convirtiéndose en uno de los primeros países del mundo en regular la IA. Las normas solo se aplican a los servicios disponibles para el público en general en China y exigen que los proveedores de servicios de IA realicen revisiones de seguridad y registren sus algoritmos ante el gobierno si sus servicios pueden influir en la opinión pública o pueden “movilizar” al público. Además, las normas obligan a revelar los contenidos generados artificialmente, así como los datos utilizados para entrenar cualquier modelo lingüístico de gran tamaño³³. En términos generales, la regulación en China viene dada por un conjunto de regulaciones departamentales que se centran en diferentes aspectos de los sistemas de IA.

Las normas generales de China y EE.UU sobre la IA en 2023 tienen algunas similitudes con la Ley Europea de Inteligencia Artificial, como la preocupación por la seguridad, la transparencia y la rendición de cuentas de los sistemas de IA, o la necesidad de revelar los contenidos generados artificialmente. Sin embargo, también hay diferencias significativas, como el ámbito de aplicación de las normas, el grado de intervención y control del gobierno, o el enfoque basado en los derechos humanos y los valores democráticos que caracteriza a la regulación europea de inteligencia artificial.

³² Ver Diario La Ley. <https://diariolaley.laleynext.es/dll/2023/10/31/el-presidente-biden-emite-una-orden-ejecutiva-para-desarrollar-en-los-eeuu-una-inteligencia-artificial-segura-y-confiable>

³³ Ver Diario La Ley. <https://diariolaley.laleynext.es/dll/2023/09/01/china-aprueba-una-regulacion-de-la-inteligencia-artificial-y-de-la-inteligencia-artificial-generativa>

Al igual que su contraparte en los Estados Unidos, el gobierno chino promueve el desarrollo y la consolidación de sus empresas al facilitar su crecimiento inicial y su posición en el mercado. Esta estrategia se lleva a cabo mediante la implementación de legislación que se centra exclusivamente en la reducción de riesgos inaceptables o en la protección significativa de sus ciudadanos. No obstante, a diferencia de otras aproximaciones, el gobierno chino elige no emplear la regulación como un instrumento directivo que pueda condicionar o restringir los avances en el campo de la inteligencia artificial (Hernández Peña, J.J. 2022).

3. *Regulación en España*

En España el documento de referencia es “La carta de derechos digitales”, que fue presentada el 14 de julio de 2021. La Carta, sin carácter normativo, se configura como una guía de futuros proyectos legislativos y del desarrollo de las políticas públicas (Barrio Andrés, M. 2021). Su objetivo es reforzar los derechos de la ciudadanía, generar certidumbre en la nueva realidad digital y aumentar la confianza ante los cambios y disrupciones tecnológicas (Barranco Avilés, MC. Y Martínez García, 2021.). En resumen, la Carta recopila principios y derechos que orientarán la formulación de futuras normativas y políticas públicas, con el propósito de garantizar la protección de los derechos individuales y colectivos en los nuevos entornos digitales. (Cotino Hueso, L. 2022)

La Carta de Derechos Digitales aborda cuestiones como la pseudonimidad (Adsuares Varela B. 2022) y la protección de datos, el derecho a la herencia digital, a la libertad de expresión y de acceso a información veraz, a la desconexión digital, la protección de menores, la ciberseguridad, la educación digital, la accesibilidad y neutralidad de internet, el uso de mecanismos de localización o perfilado, el acceso a los archivos, o el derecho de cada persona al control sobre su propia identidad así como el derecho a la igualdad y no discriminación, los derechos de la persona ante la inteligencia artificial y las neurotecnologías

Por último, resulta de interés el Real Decreto 22/2023, por el que se aprueba el estatuto de la Agencia Española de Supervisión de la Inteligencia Artificial (AESIA). La AESIA se adscribe al Ministerio de Asuntos Económicos y Transformación Digital a través de la Secretaría de Estado de Digitalización e Inteligencia Artificial. Con la creación de esta Agencia, España se convierte en el primer país europeo en tener un órgano de estas características y se anticipa a la entrada en vigor del Reglamento Europeo de Inteligencia Artificial. Dicho proyecto de reglamento establece para los Estados miembros la obligación de seleccionar una autoridad nacional de supervisión

que se encargue de supervisar la aplicación de la normativa en materia de Inteligencia Artificial³⁴

IV. DECLARACIONES Y ACUERDOS INTERNACIONALES BASADOS EN LA ÉTICA

La Recomendación sobre la ética de la inteligencia artificial, de la UNESCO adoptada el 23 de noviembre de 2021³⁵ representa un hito significativo en el ámbito de la ética de la inteligencia artificial, al convertirse en el primer marco normativo universal sobre esta cuestión, y que tiene como objetivo orientar a los Estados, las organizaciones internacionales, el sector privado, la sociedad civil y la comunidad científica en el desarrollo y el uso de la IA, de acuerdo con los valores universales de los derechos humanos, la dignidad humana y la diversidad cultural³⁶ La Recomendación defiende ciertos valores y descansa en determinados principios, referidos a distintos ámbitos de actuación política. Aborda la ética de la IA como una

“reflexión normativa sistemática, basada en un marco integral, global, multicultural y evolutivo de valores, principios y acciones interdependientes, que puede guiar a las sociedades a la hora de afrontar de manera responsable los efectos conocidos y desconocidos de las tecnologías de la IA en los seres humanos, las sociedades y el medio ambiente y los ecosistemas, y les ofrece una base para aceptar o rechazar las tecnologías de la IA”.

El enfoque central de esta declaración radica en la promoción y salvaguardia de los derechos humanos, la dignidad humana y la sostenibilidad medioambiental. Además, se enfatiza la importancia de principios fundamentales como la rendición de cuentas y el Estado de derecho. Asimismo,

³⁴ Real Decreto 729/2023, de 22 de agosto, por el que se aprueba el Estatuto de la Agencia Española de Supervisión de Inteligencia Artificial. https://www.boe.es/diario_boe/txt.php?id=BOE-A-2023-18911; Vid. doc. Referencia de Consejo de Ministros Aprobado el estatuto de la Agencia Española de Supervisión de la Inteligencia Artificial <https://www.hacienda.gob.es/Documentacion/Publico/GabineteMinistro/Notas%20Prensa/2023/CONSEJO-DE-MINISTROS/22-08-23-NP-CM-Estatutos-Agencia-Inteligencia-Artificial.pdf>

³⁵ Fue respaldada por unanimidad por los 193 Estados Miembros de la UNESCO en noviembre de 2021

³⁶ Adopción del primer instrumento normativo mundial sobre la ética de la inteligencia artificial. En la Conferencia General de la UNESCO celebrada en noviembre de 2021, los 193 Estados Miembros de la Organización adoptaron una Recomendación sobre la Ética de la Inteligencia Artificial, el primer instrumento normativo existente a nivel mundial sobre esta cuestión.

<https://www.unesco.org/es/articles/adopcion-del-primer-instrumento-normativo-mundial-sobre-la-etica-de-la-inteligencia-artificial>

esta declaración abarca capítulos específicos que abogan por una gobernanza más efectiva de los datos, así como por la inclusión y la igualdad de género. La aplicación de la ética a la IA debe estar presente para cualquier persona, sea de naturaleza pública o privada, que participe en cualquier etapa del ciclo de vida de un sistema de IA, incluyendo la investigación, concepción y el desarrollo y su utilización, además del mantenimiento, el funcionamiento, la comercialización, la financiación, el seguimiento y la evaluación, la validación, el fin de la utilización, el desmontaje y la terminación³⁷. La Recomendación presta especial atención a las repercusiones éticas más amplias de los sistemas de IA en las principales esferas de competencia de la UNESCO. En concreto, en la educación, la ciencia, la cultura y la comunicación y la información.

Además, es preciso hacer referencia a los acuerdos internacionales que se han tomado recientemente sobre esta materia.

La declaración de Bletchley Park es un hito histórico en la cooperación internacional sobre la inteligencia artificial (IA), que reconoce los beneficios y los riesgos que plantea esta tecnología para la humanidad y el planeta. La declaración fue firmada por los representantes de 28 países, entre ellos la Unión Europea, Estados Unidos y China, que son los principales actores en el desarrollo y el uso de la IA. La declaración se basa en los principios y valores compartidos de la democracia, el Estado de Derecho, los derechos humanos y la sostenibilidad, y se compromete a trabajar conjuntamente para garantizar una IA segura, centrada en el ser humano, confiable y responsable.

La declaración de *Bletchley Park* es el primer paso de un proceso continuo de cooperación global sobre la IA, que tendrá continuidad con futuras cumbres cada seis meses en diferentes países. La próxima cumbre tendrá lugar en Corea del Sur, y después en Francia. El propósito de estas cumbres es crear un panel al estilo del IPCC del cambio climático, que evalúe periódicamente los avances y amenazas de la IA, y que emita recomendaciones y

³⁷ Resolución del Parlamento Europeo, de 20 de octubre de 2020, sobre un marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas, 2020/2012(INL).

Comisión Europea, *Generar confianza en la inteligencia artificial centrada en el ser humano* [COM(2019) 168].; Grupo de expertos de alto nivel sobre inteligencia artificial, *Directrices éticas para una IA fiable*, 2019.; Grupo de expertos de alto nivel sobre inteligencia artificial, *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment (Lista de evaluación para una inteligencia artificial fiable con fines de autoevaluación)*, 2020.; La Alianza de la IA es un foro de múltiples interesados que se creó en junio de 2018. Para más información, véase el siguiente enlace: <https://ec.europa.eu/digital-single-market/en/european-ai-alliance>.; Grupo de expertos de alto nivel sobre IA, *Directrices éticas para una IA fiable*, 2019.

orientaciones para los gobiernos, las organizaciones internacionales, el sector privado, la sociedad civil y la comunidad científica³⁸.

Por último, el Grupo de los Siete (G7) puso en marcha el pasado mayo de 2023 en la cumbre de Hiroshima (oeste) una iniciativa que ha derivado en la elaboración de once principios para empresas y organizaciones involucradas en el desarrollo y el uso de sistemas de IA. Estos principios se basan en los valores compartidos por el G7, como la democracia, el estado de derecho, el respeto a los derechos humanos y la diversidad.

El código de conducta se denomina el “Proceso de IA de Hiroshima” y tiene como objetivo promover el desarrollo de sistemas de IA seguros y fiables a nivel internacional y gestionar sus riesgos. El código de conducta se revisará periódicamente para adaptarse a la rápida evolución de la tecnología y se cooperará con los actores implicados en el sector de la IA, como la sociedad civil, la academia, el sector privado y las organizaciones internacionales. El código de conducta fue respaldado por los líderes del G7 (Alemania, Canadá, Estados Unidos, Francia, Italia, Japón y el Reino Unido) el 30 de octubre de 2023.

V. LOS DERECHOS FUNDAMENTALES FRENTE A LA INTELIGENCIA ARTIFICIAL

El uso de la inteligencia artificial, con sus características específicas, como son la utilización masiva de datos, la falta de transparencia, o su comportamiento autónomo, puede tener consecuencias negativas para diversos derechos fundamentales, a partir de una posible afectación a la dignidad humana que se plantea como base y punto de referencia del conjunto de los derechos fundamentales³⁹. Pese a que los derechos no se estudian de forma aislada, en cuanto componen un sistema, un conjunto, que conlleva que solo se pueden entender dentro de ese sistema. A los meros efectos de una mejor exposición, resulta conveniente estructurar en cuatro apartados la referencia sobre cómo afectan los sistemas IA al régimen de derechos fundamentales.

El primero, el principio de igualdad, como consecuencia de las cuestiones relacionadas con la discriminación algorítmica y la necesidad de garantizar la igualdad y la no discriminación en el uso de sistemas de IA. Lo que se conoce

³⁸ Declaración de Bletchley de los países que asisten a la Cumbre de seguridad de la IA, 1 y 2 de noviembre de 2023 <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/dbc58681-1b68-47e0-8e3f-f91435fdf8ce>

³⁹ Con la referencia al derecho a la dignidad (art. 10 CE) fundamento del orden político y la paz social: La dignidad de la persona; Los derechos inviolables que le son inherentes; El libre desarrollo de la personalidad; El respeto a la ley y a los derechos de los demás.

como sesgo algorítmico (Belloso Martín, N. 2022). El sesgo puede producirse de forma intencional o de manera involuntaria. De manera que determinados puntos se programen en un algoritmo, o no lo hagan, con lo que se consigue replicar sesgos de carácter estructural. O bien, la IA podría ser utilizada para la toma de decisiones que están afectadas por cuestiones como pueden ser el sexo o la pertenencia a un grupo diferenciado por edad o por origen social, u otros.

Los sesgos son especialmente problemáticos cuando conducen a discriminaciones a favor de un grupo en detrimento de otro. Este fenómeno no es exclusivo de los sistemas de inteligencia artificial, sino que es inherente a cualquier proceso de toma de decisiones, ya sea ejecutado por seres humanos o de forma automática. Es crucial destacar que estos sesgos no se limitan a simples errores técnicos; sus implicaciones son significativas. Se ha observado que tienden a repetirse y a perpetuarse, fortaleciendo dinámicas de dominación, privilegio y discriminación. En última instancia, esto amplifica el riesgo de que las desigualdades existentes se vean acentuadas y consolidadas a través de la automatización y el aprendizaje de las máquinas (Ninareh Mehrabi, N.; Morstatter, F.; Saxena, N.; Lerman, K.; Galstyan, A. 2019).

La segunda gran cuestión, relacionada con la anterior, pero que puede analizarse desde un enfoque más específico y pone en relación los sistemas IA con los derechos fundamentales, es la protección de la vida privada. En el ámbito del artículo 18 de la Constitución, aparecen nuevos riesgos o, en algunos casos los mismos pero llevados a una dimensión distinta y mucho mayor como consecuencia de la IA. La recolección y el análisis masivo de datos plantean interrogantes sobre el alcance de la privacidad individual y la salvaguarda de la intimidad en un entorno de creciente vigilancia y monitoreo (Cotino Hueso, L. 2017)⁴⁰.

La fuente de los datos utilizados por la inteligencia artificial es una consideración crítica que plantea diversas preguntas en el ámbito jurídico. ¿De dónde provienen estos datos? ¿Fueron obtenidos de manera ética y respetando los derechos de privacidad, o existe la posibilidad de que se hayan recopilado de manera que infrinja dichos derechos? Además, surge la preocupación sobre la autenticidad de los datos: ¿son legítimos o potencialmente falsos, careciendo de valor? Más aún, la representatividad de los datos es crucial; si no son representativos, los resultados pueden ser sesgados, conduciendo a discriminación arbitraria.

La procedencia y contenido de los datos, por lo tanto, son elementos fundamentales que generan una serie de interrogantes en el contexto legal relacionadas con la inteligencia artificial. Es esencial analizar si los datos

⁴⁰ AEPD – ISMS Forum (eds.), ACED, E. et al. (coords.) Código de buenas prácticas en protección de datos para proyectos de Big Data, AEPD e ISMS Forum, 2017 Madrid.

utilizados son válidos, éticos y representativos para garantizar resultados justos y no sesgados. La IA también supone riesgos para la privacidad y la protección de datos al utilizarse, por ejemplo, en equipos de reconocimiento facial o para el seguimiento en línea y la creación de perfiles de personas. Además, permite fusionar información que una persona ha proporcionado con datos nuevos, pudiendo crear perfilados (Andoni Eguíluz Castañeira J. 2020). Los sistemas de IA también pueden usarse para crear vídeos, audios o imágenes falsos pero realistas, conocidos como “*deepfakes*”. Este contenido puede implicar riesgos financieros, daños reputacionales y problemas en las tomas decisiones. Todo esto podría conducir a la separación y polarización en la esfera pública y manipular las elecciones.

Finalmente, respecto de los aspectos más reseñables que ponen en relación el Derecho a la protección de datos personales con los sistemas IA (Rebollo Delgado L. 2023) El RGPD prevé una especial limitación aplicable a los sistemas de IA y referida a la toma de decisiones automatizadas. Así, las personas tienen el derecho de no ser sujetas a decisiones automatizadas que, basadas únicamente en el procesamiento de datos, puedan tener un impacto legal significativo en ellas o afectarlas de manera similar⁴¹.

Todo interesado tendrá derecho a no ser objeto de una decisión basada únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzca efectos jurídicos en él o le afecte significativamente de modo similar. (art. 22 RGPD)

Este derecho garantiza que ninguna persona sea sujeta a decisiones importantes o significativas que afecten sus derechos y su vida diaria cuando estas decisiones se basan únicamente en procesos automatizados, mediante algoritmos o sistemas informáticos, sin intervención humana directa (Palma Ortigosa A. 2022.) Así, la negación de un préstamo, una oferta de trabajo, la aprobación o rechazo de una solicitud de seguro o la asignación de recursos públicos no debe basarse únicamente en algoritmos sin que una persona haya tenido la oportunidad de influir en la decisión.

El tratamiento automatizado de datos también abarca la elaboración de perfiles⁴², que es la evaluación de datos personales para analizar o predecir aspectos relacionados con el desempeño laboral, la situación económica, la

⁴¹ “Art. 22 del RGPD

⁴² Debe ser una forma automatizada de tratamiento, incluyen do aquellos tratamientos que tienen participación parcialmente humana. Debe llevarse a cabo respecto a datos personales; Y el objetivo de la elaboración de perfiles debe ser evaluar aspectos personales sobre una persona física. Vid. Informe de la agencia de Protección de Datos sobre adecuación al Reglamento General de Protección de Datos de tratamientos que incorporan inteligencia artificial. Una introducción

<https://www.aepd.es/documento/adecuacion-rgpd-ia.pdf>

salud, preferencias personales, fiabilidad, comportamiento, ubicación o movimientos de una persona, cuando estas evaluaciones pueden tener consecuencias legales significativas o un impacto similar⁴³.

“Toda forma de tratamiento automatizado de datos personales consistente en utilizar datos personales para evaluar determinados aspectos personales de una persona física, en particular para analizar o predecir aspectos relativos al rendimiento profesional, situación económica, salud, preferencias personales, intereses, fiabilidad, comportamiento, ubicación o movimientos de dicha persona física”. (Art. 4.4 RGPD)

Teóricamente son cuestiones distintas. Así, para una plataforma que utiliza tratamiento automatizado, el sistema podría, por ejemplo, registrar información, asignar un código de seguimiento a esta información y ofrecer respuestas automáticas para guiar a un usuario, o tomar decisiones en un determinado proceso. En cuanto a la elaboración de perfiles, la plataforma lo que haría es analizar datos y patrones de comportamiento para obtener una información gracias a la que podría, por ejemplo, personalizar recomendaciones.

La creación de perfiles tiene el potencial de perpetuar los estigmas preexistentes y fomentar la fragmentación social. De manera paralela, puede encajonar a un individuo en una etiqueta específica, restringiendo sus opciones a las sugerencias preestablecidas. Este fenómeno puede minar la libertad de elección de las personas, limitando, por ejemplo, sus decisiones sobre productos y servicios, como libros, música o noticias. En ocasiones, la creación de perfiles puede conllevar predicciones erróneas, mientras que en otros casos puede dar lugar a la negación de servicios y bienes, perpetrando así una discriminación injustificada.

El Reglamento establece disposiciones con el propósito de garantizar que la práctica de elaboración de perfiles y la toma de decisiones automatizadas, ya sea con o sin la elaboración de perfiles, se realicen de manera que no conlleven un impacto injustificado en los derechos de los individuos. Entre estas disposiciones se incluyen: La imposición de requisitos específicos en cuanto a transparencia y equidad; La asignación de mayores responsabilidades en términos de proactividad; La definición de bases legales específicas para el procesamiento de datos; El otorgamiento de derechos individuales para oponerse a la elaboración de perfiles, en particular con fines de marketing; La exigencia de llevar a cabo una evaluación de impacto en la protección de datos cuando se cumplan ciertas condiciones.

⁴³ Los campos del sector bancario y financiero, la asistencia sanitaria, la fiscalidad, los seguros, la mercadotecnia y la publicidad son ejemplos típicos en los cuales se lleva a cabo de manera recurrente la elaboración de perfiles para enriquecer el proceso de toma de decisiones.

La elaboración de perfiles está formada por tres elementos: debe ser una forma automatizada de tratamiento, debe llevarse a cabo respecto a datos personales y el objetivo de la elaboración de perfiles debe ser evaluar aspectos personales sobre una persona física⁴⁴. Implica un proceso de tratamiento automatizado, aunque no excluye la participación humana en la definición de dichos perfiles. Este procedimiento se basa en deducciones estadísticas y se utiliza comúnmente para predecir comportamientos o características de individuos a partir de datos recolectados de diversas fuentes, considerando similitudes estadísticas con otros individuos.

Según el Reglamento General de Protección de Datos (RGPD), la creación de perfiles se define como el procesamiento automatizado de datos personales con el fin de evaluar aspectos personales, especialmente para analizar o prever aspectos de las personas. La palabra “evaluar” implica que la creación de perfiles involucra algún tipo de valoración o juicio sobre un individuo. Sin embargo, es importante destacar que la mera clasificación de personas en función de datos conocidos como edad, sexo y altura no siempre constituye una creación de perfiles, ya que esto dependerá de la finalidad de dicha clasificación. Por ejemplo, una empresa podría clasificar a sus clientes en función de su edad o género con fines estadísticos y para obtener una visión general de su base de clientes, sin realizar predicciones o sacar conclusiones sobre un individuo en particular. En este caso, la finalidad no es evaluar las características individuales, por lo que no se consideraría una creación de perfiles⁴⁵.

En textos anteriores, en los que se inspira el RGPD, se recoge que la elaboración de perfiles puede implicar tres fases distintas: Primera, recogida de datos; Segunda: análisis automatizado para identificar correlaciones; Tercera: aplicación de la correlación a una persona para identificar características de comportamientos presentes o futuros. Los responsables que llevan a cabo la elaboración de perfiles deberán garantizar que cumplen los requisitos del RGPD respecto a todas estas fases⁴⁶. La elaboración de perfiles, en líneas generales, implica la recopilación de información relativa a una persona o un grupo de individuos, seguida de una evaluación de sus características y

⁴⁴ Grupo de Trabajo sobre Protección de Datos del artículo 29. Directrices del Comité Europeo sobre Protección de Datos referentes a decisiones individuales automatizadas y elaboración de perfiles a los efectos del reglamento 2016/679. Adoptadas el 3 de octubre de 2017 y revisadas y adoptadas por última vez el 6 de febrero de 2018

<https://www.aepd.es/documento/wp251rev01-es.pdf>

⁴⁵ *Ibid.*

⁴⁶ La Recomendación CM/Rec(2010)13 del Comité de Ministros a los Estados miembros sobre la protección de las personas con respecto al tratamiento automatizado de datos de carácter personal en el contexto de la creación de perfiles

https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=09000016805cdd2a

patrones de comportamiento con el fin de categorizarlos de manera específica. Esto se realiza principalmente para llevar a cabo análisis o pronósticos acerca de aspectos como su competencia en la realización de tareas, sus intereses personales o su comportamiento futuro. La determinación de si se configura una decisión automatizada o no, se hace depender de las circunstancias particulares involucradas en el proceso⁴⁷.

En tercer lugar, los riesgos de la IA planean sobre los derechos del ámbito de la comunicación. Resulta especialmente preocupante los daños que pueden producirse respecto de otro derecho fundamental, que queda necesariamente afectado, la libertad de expresión, más concretamente la libertad de información, que se manifiesta a través de lo que se ha denominado censura algorítmica. La censura algorítmica es el fenómeno por el cual los sistemas de inteligencia artificial (IA) limitan o manipulan la información que se difunde o se accede en el entorno digital, afectando así a la libertad de expresión y de información. Este fenómeno puede tener diversas causas y consecuencias, que se relacionan con el diseño, el funcionamiento y el uso de los sistemas de IA. (Teruel Lozano G. 2014; Cotino Hueso, L, 2005, p. 59).

La inteligencia artificial tiene, además, la capacidad de difundir información errónea; desde otra perspectiva, puede contribuir a la desinformación y la propagación de noticias falsas. Asimismo, la inteligencia artificial puede generar artificialmente imágenes, documentos y voces, convirtiéndose en un medio que puede afectar de manera significativa bienes jurídicos fundamentales. Estos dos aspectos adicionales subrayan la importancia de considerar la influencia potencialmente perjudicial de la inteligencia artificial en la calidad de la información y en la integridad de diversos elementos, destacando así la necesidad de un enfoque crítico y regulaciones adecuadas en estos ámbitos.

Lo anterior se relaciona también con el derecho de sufragio como base del proceso democrático (González de la Garza, M. 2019). Se ha señalado la necesidad de preservar este derecho en la era digital. La desinformación y los ataques informáticos dirigidos contra procesos electorales, campañas y

⁴⁷ Se pueden tener en cuenta una serie de aspectos para ponderar sobre la necesidad de los intereses legítimos por el responsable del tratamiento o por un tercero: *el nivel de detalle del perfil (un interesado cuyo perfil corresponde a un grupo ampliamente descrito como «gente interesada en la literatura inglesa», o segmentado y abordado a un nivel muy específico); la exhaustividad del perfil (si el perfil describe únicamente un pequeño aspecto del interesado, o presenta una visión más exhaustiva); las consecuencias de la elaboración del perfil (los efectos sobre el interesado); y la garantías destinadas a garantizar la lealtad, la no discriminación y la precisión del proceso de elaboración de perfiles.* Grupo de Trabajo sobre Protección de Datos del artículo 29. Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del reglamento 2016/679. Adoptadas el 3 de octubre de 2017 y revisadas y adoptadas por última vez el 6 de febrero de 2018 <https://www.aepd.es/documento/wp251rev01-es.pdf>

candidatos pueden polarizar el discurso público y quebrantar el secreto del voto, socavar la integridad y equidad del proceso electoral y minar la confianza de la ciudadanía en los representantes que hayan sido elegidos. En este contexto, se ha destacado la importancia de las salvaguardias y las medidas activas para contrarrestar las campañas de desinformación, el uso indebido de datos privados, las amenazas híbridas y los ciberataques (Abacatoria, A. 2019)⁴⁸.

Para garantizar el respeto a estos derechos en la era digital, se requiere un marco normativo adecuado, que regule el uso de la IA de forma ética y responsable, y que establezca mecanismos de transparencia, rendición de cuentas y control de los sistemas de IA. Asimismo, se requiere una mayor concienciación y educación de los ciudadanos, que les permita ejercer un pensamiento crítico y una alfabetización mediática e informacional, frente a los riesgos de la censura algorítmica.

Por último, se hace referencia al derecho a la tutela judicial efectiva, reconocido en el artículo 24 de la Constitución Española, que puede verse limitado por la falta de vigilancia humana cuando se utilizan sistemas de inteligencia artificial (IA) para asistir o sustituir a los jueces en la aplicación de la ley. La IA puede ofrecer ventajas como la rapidez, la eficiencia o la uniformidad, pero también puede presentar riesgos como el error, el sesgo o la falta de transparencia. Por ello, es necesario que exista un control humano adecuado sobre la IA, que garantice el respeto a los principios y garantías procesales, así como a los derechos y libertades de las personas (Barona Villar, S. 2021)

Resulta necesario establecer unos criterios de calidad, seguridad y ética para la IA, que aseguren su fiabilidad, precisión y equidad, como forma de ayudar a preservar el derecho a la tutela judicial efectiva, así como manera de fomentar la confianza y la aceptación de la IA en el ámbito judicial. El control o vigilancia humano facilitará el respeto de éste y otros derechos fundamentales, ya que contribuye a reducir al mínimo el riesgo de adoptar decisiones asistidas por IA erróneas o sesgadas de aplicación de la ley por el poder judicial (Lorenzo Pérez, C. 2022).

VI. CONCLUSIONES

La regulación efectiva de la IA requiere una comprensión común de su significado. Se ha buscado definir los sistemas IA según sus capacidades y a través de sus características. Principalmente, las cuestiones referentes a la

⁴⁸ La Carta de los Derechos Fundamentales en el contexto de la inteligencia artificial y el cambio digital <https://data.consilium.europa.eu/doc/document/ST-11481-2020-INIT/es/pdf>

explicabilidad y autonomía. La *explicabilidad* implica comprender y comunicar los resultados de la IA, especialmente en modelos complejos como el *deep learning*. La autonomía, que abarca desde ejecutar instrucciones preprogramadas hasta la toma de decisiones independiente, plantea preocupaciones sobre la capacidad de la IA para causar daño físico y la asignación de responsabilidad. En este contexto, se propone un enfoque basado en las capacidades funcionales de la IA para garantizar neutralidad tecnológica y adaptabilidad regulatoria ante futuros avances.

Del análisis del marco normativo existente se traducen las dificultades que se presentan para su armonización global. Se ha prestado especial atención al Reglamento de la Unión Europea y al posterior acuerdo de 8 de diciembre, así como al resto de los documentos y propuestas de la Unión, que evidencian la preocupación por dotar de un marco regulador a los derechos fundamentales ante la irrupción de los sistemas de IA. Estos derechos pueden verse afectados por los sesgos, las discriminaciones, la desinformación y las amenazas a la ciberseguridad que puede generar la IA, así como por la falta de control y responsabilidad de sus acciones y decisiones.

La ética y los acuerdos globales resultan fundamentales para dotar de principios que garanticen los derechos fundamentales ante la irrupción de los sistemas IA. La Recomendación sobre la ética de la inteligencia artificial, de la UNESCO, se convierte en el primer marco normativo universal sobre esta cuestión, que establece los valores, principios y ámbitos de actuación política que deben orientar a los actores involucrados en el ciclo de vida de un sistema de IA. Asimismo, se ha hecho referencia a otros acuerdos internacionales relacionados con la ética de la IA, como la declaración de Bletchley Park, que es una iniciativa de cooperación global entre 28 países, y los once principios del G7 para empresas y organizaciones que trabajan con sistemas de IA. Estos acuerdos señalan la necesidad de trabajar en marcos regulatorios globales y comunes, que respeten los derechos humanos, la dignidad humana y la diversidad cultural.

Definida la inteligencia artificial (IA) y su regulación, se plantea cómo puede afectar a los derechos fundamentales, especialmente al principio de igualdad, a la protección de la vida privada, a la libertad de expresión y a la participación democrática. Así como a la toma de decisiones por el poder judicial. El hecho de que la IA pueda tener un comportamiento autónomo y poco transparente dificulta especialmente el control y la responsabilidad de sus acciones y decisiones.

El derecho a la protección de la vida privada se ve amenazado por el uso masivo de datos personales por parte de los sistemas de IA. Se plantean diversas cuestiones sobre la procedencia, el contenido, la autenticidad y la representatividad de los datos utilizados por la IA, así como sobre los riesgos de sesgos, discriminaciones, desinformación y ciberataques que puede

generar la IA. Se evidencian también las posibles consecuencias negativas de la IA para la esfera pública, la democracia y las elecciones, por lo que se proponen medidas de control y educación para garantizar el respeto a los derechos fundamentales. Igualmente, es preciso destacar los riesgos de decisiones automatizadas que produzcan efectos jurídicos o afecten significativamente a las personas, así como de la elaboración de perfiles.

Finalmente, se ha profundizado en el impacto de la IA en los derechos de carácter político, relacionados con el ámbito de la comunicación, como la libertad de expresión, la libertad de información y el derecho de sufragio. Se ha descrito el fenómeno de la censura algorítmica, así como los riesgos de que la IA puede difundir información errónea, generar información falsa y atentar contra la integridad y equidad de los procesos electorales. Se plantea la necesidad de un marco normativo, un control humano y una educación ciudadana para garantizar el respeto a estos derechos en la era digital. Por último, se ha abordado el derecho a la tutela judicial efectiva, que puede verse limitado por el uso de la IA para asistir o sustituir a los jueces. Se ha propuesto establecer criterios de calidad, seguridad y ética para la IA, que aseguren su fiabilidad, precisión y equidad, así como un control humano adecuado que reduzca el riesgo de decisiones erróneas o sesgadas.

VII. REFERENCIAS BIBLIOGRÁFICAS

- ABA CATORIA, Ana. “Seguridad nacional: libertad y seguridad en el ciberespacio”. en *RGDA Iustel*, n.º 50, febrero 2019.
- ADSUARA VARELA, Borja. “El derecho al pseudonimato. Entre la identificación y el anonimato” Lorenzo de Cotino Hueso (editor), *La Carta de Derechos Digitales*, Tirant Lo Blanch, Valencia, 2022, pp. 53 a 78.
- ARELLANO TOLEDO, Wima. “El derecho a la transparencia algorítmica en big data e inteligencia artificial”, en *RGDA Iustel*, n.º 50, febrero 2019.
- BALAGUER CALLEJÓN, Francisco. *La constitución del algoritmo* Fundación Giménez Abad. Zaragoza, 2023. 2º ed.
- BARONA VILLAR, Silvia. *Algoritmización del derecho y de la justicia. De la inteligencia artificial a la Smart Justice*. Tirant lo Blanch, Valencia, 2021.
- BARRIO ANDRÉS, Moisés. *Delitos 2.0. Aspectos penales, procesales y de seguridad de los ciberdelitos*. Madrid, Wolters Kluwer, 2018.
- “Carta de Derechos Digitales: ¿el preámbulo de una reforma de la Constitución española?, *The conversation* (2021).
- Génesis y desarrollo de los derechos digitales. *Revista De Las Cortes Generales*, (110), 197-233. <https://doi.org/10.33426/rcg/2021/110/1572>.
- *Manual de Derecho digital*, Editorial Tirant lo Blanch, 2.ª edición (2022).
- BARRANCO AVILÉS, María del Carmen y MARTÍNEZ GARCÍA, María. “La Carta de Derechos Digitales: una oportunidad para el desarrollo de una inteligencia artificial ética y responsable” *Revista Icade*, n.º 116, 2021.

- BELLOSO MARTÍN, Nuria. “La problemática de los sesgos algorítmicos (con especial referencia a los de género). ¿Hacia un derecho a la protección contra los sesgos?” En oc. Dir. Fernando Llano Alonso *Inteligencia artificial y filosofía del Derecho*. Ediciones Laborum. Murcia, 2022, pp. 45-78.
- BOIX PALOP, Andrés. “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*, n.º 1, 2020. Sobre la regulación de los algoritmos en Derecho administrativo.
- COTINO HUESO, Lorenzo y CASTELLANOS CLARAMUNT, Jorge. “Transparencia y Explicabilidad de la Inteligencia Artificial y “Compañía” (Comunicación, Interpretabilidad, Intelligibilidad, Auditabilidad, Testabilidad, Comprobabilidad, Simulabilidad...). Para qué, para quién y cuánta”. En *Transparencia y explicabilidad de la Inteligencia Artificial*. Tirant Lo Blanch. Valencia 2022.
- COTINO HUESO, Lorenzo (editor). *La Carta de Derechos Digitales*, Tirant Lo Blanch, Valencia, 2022.
- . “Big data e inteligencia artificial, una aproximación a su tratamiento jurídico desde los derechos fundamentales”, en: *Dilemata*, 24, 2017 pp. 131-150.
- . “Un análisis crítico constructivo de la Propuesta de Reglamento de la Unión Europea por el que se establecen normas armonizadas sobre la Inteligencia Artificial (Artificial Intelligence Act)”, *Diario La Ley*. 2021.
- . “Algunas Claves para el análisis constitucional futuro de las libertades públicas ante las nuevas tecnologías (con especial atención al fenómeno de los «blogs»)”, en *Estudios jurídicos sobre la sociedad de la información y nuevas tecnologías*, Facultad de Derecho de Burgos, Burgos, 2005, p. 59.
- EGÚILUZ CASTAÑEIRA, Josu Andoni. “Desafíos y retos que plantean las decisiones automatizadas y los perfiles para los derechos fundamentales”, en: *Estudios de Deusto*, 68, 2020, 325-368.
- GONZÁLEZ DE LA GARZA, Miguel. “El Big Data como la gran herramienta de pronóstico de tendencias en los sistemas electorales. del salto del paradigma demoscópico a los datos de comportamiento real”. en *RGDA Iustel*, n.º 50, febrero 2019.
- GUTIÉRREZ DAVID, María Estrella. “Administraciones inteligentes y acceso al código fuente y los algoritmos públicos. Conjorando riesgos de cajas negras decisionales”, en *Derecom*, n.º 31, pp. 19-105, ver pp. 55-56, 2021, <http://www.derecom.com/derecom/>
- HERNÁNDEZ PEÑA, Juan Carlos. *El marco jurídico de la inteligencia artificial. Principios, procedimientos y estructura de gobernanza* Aranzadi. 2022.
- LORENZO PÉREZ, Cristina. *Inteligencia artificial en la administración de Justicia: regulación española y marco europeo e internacional*. Proyectos desarrollados por el Ministerio de Justicia de España. Dirección General de Transformación Digital de la Administración de Justicia. Ministerio de Justicia. 2022. <https://www.cej-mjusticia.es/sede/publicaciones/ver/13637>
- MARTÍN DELGADO, Isaac. “La gestión electrónica del procedimiento administrativo”, *QDL* 21, octubre de 2009 y, del mismo autor, “Naturaleza, concepto y régimen jurídico de la actuación administrativa automatizada”, *Revista de Administración Pública*, n.º 180, Madrid, septiembre-diciembre 2009.

- MEHRABI, Ninareh; MORSTATTER, Fred; SAXENA, Nripsuta; LERMAN, Kristina; GALSTYAN, Aram. "A Survey on Bias and Fairness in Machine Learning." (2019) <https://doi.org/10.48550/arXiv.1908.09635>
- MERCHÁN MURILLO, Antonio. "Inteligencia artificial y Blockchain: retos jurídicos en paralelo" En Análisis monográfico. Derecho Público, derechos y transparencia ante el uso de algoritmos, inteligencia artificial y big data (Coordinado por Andrés Boix Palop y Lorenzo Cotino Hueso). *Revista General de Derecho Administrativo*. 2019.
- PALMA ORTIGOSA, Adrián. *Decisiones automatizadas y protección de datos personales. Especial atención a los sistemas de inteligencia artificial*. Dykinson, 2022.
- REBOLLO DELGADO, Lucrecio. *Inteligencia artificial y Derechos fundamentales*. Madrid: Dykinson, 2023.
- SAMEK, Wojciech; MONTAVON, Grégoire; VEDALDI, Andrea; HANSEN, Lars Kai, MÜLLER, Klaus-Robert. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*: 11700 (Lecture Notes in Artificial Intelligence) Springer. 2019.
- TERUEL LOZANO, Germán. "Libertad de expresión y censura en internet", *Estudios De Deusto* 62 (2), 41-72. [https://doi.org/10.18543/ed-62\(2\)-2014pp41-72](https://doi.org/10.18543/ed-62(2)-2014pp41-72).
- VESTRI, Gabriele. "La inteligencia artificial ante el desafío de la transparencia algorítmica: Una aproximación desde la perspectiva jurídico-administrativa", *Revista Aragonesa de Administración Pública*, n.º 56, Zaragoza, 2021, pp. 368-398.