FACULTAD DE DERECHO

INTELIGENCIA ARTIFICIAL Y DERECHO

Un análisis sobre el impacto en el Estado de Derecho

Autor: Gonzalo Gutiérrez Evans

5ºE-3 B Doble Grado Derecho y ADE

Filosofía del Derecho

Tutor: Prof. José María Lasalle Ruiz

Madrid

Junio 2024

RESUMEN

Los avances en la ciencia y los cambios sociales han ido históricamente de la mano, suponiendo las grandes revoluciones tecnológicas autenticas reestructuraciones de las sociedades humanas. Siendo el Derecho el principal agente regulador de la vida en sociedad, no es de extrañar que éste y el progreso tecnológico estén estrechamente vinculados, sobre todo si se tiene en cuenta el papel que debe jugar la ley en el Estado de Derecho al establecer los límites y reglas que rigen el desarrollo científico.

En este siglo XXI caracterizado por la omnipresencia de la tecnología en una sociedad global, llama especial atención el desarrollo del campo de investigación de la Inteligencia Artificial. Debido a las diversas maneras en las que puede implementarse la Inteligencia Artificial en la sociedad, y a las cuestiones éticas y filosóficas que su propia conceptualización plantea, se emprende esta investigación con el objetivo de estudiar las maneras en las que el desarrollo de las tecnologías de Inteligencia Artificial puede afectar a la configuración actual del Estado de Derecho, así como los mecanismos éticos y normativos que este tendrá que desarrollar para asegurar que la implantación de estas herramientas no perjudica a los ciudadanos en el ejercicio de sus derechos y libertades.

Palabras clave: Inteligencia Artificial, Estado de Derecho, Ética del Derecho, Gobernanza Tecnológica, Derechos Digitales.

Abstract

Advances in science and social changes have historically gone hand in hand, with major technological revolutions leading to significant restructuring of human societies. As the primary regulatory agent of societal life, it is not surprising that law and technological progress are closely linked, especially considering the role that law must play in the rule of law by establishing the limits and rules that govern scientific development. In this 21st century, characterized by the omnipresence of technology in a global society, special attention is drawn to the development of the field of Artificial Intelligence research. Due to the various ways in which Artificial Intelligence can be implemented in society, and the ethical and philosophical questions its very conceptualization raises, this research is undertaken with the objective of studying the ways in which the development of Artificial Intelligence technologies can affect the current configuration of the rule of law, as well as the ethical and regulatory mechanisms that will need to be developed to ensure that the implementation of these tools does not harm citizens in the exercise of their rights and freedoms.

Keywords: Artificial Intelligence, Rule of Law, Legal Ethics, Technological Governance, Digital Rights

ÍNDICE

LISTADO DE ABREVIATURAS

Inteligencia Artificial (IA)

Reglamento General de Protección de Datos (RGPD)

Artificial Intelligence Act (AIA)

Agencia Española de Supervisión de Inteligencia Artificial (AESIA)

Unión Europea (UE)

Estrategia Nacional de Inteligencia Artificial (ENIA)

CAPÍTULO I. INTRODUCCIÓN

Sin duda alguna, la humanidad se encuentra ahora mismo a punto de comenzar una nueva etapa en su historia, la Era de la Inteligencia Artificial. La omnipresencia de la inteligencia artificial en nuestros días es innegable, tanto en el ámbito del ocio como en el laboral y social. Aparece en nuestras conversaciones y noticiarios, domina la escena en el cine y en la literatura de ciencia ficción, y sus posibles aplicaciones en campos como la medicina, la administración de justicia o la armamentística suscitan complejos debates.

Santiago Ramón y Cajal dijo hace más de 100 años que, "*casi todos los males de pueblos e individuos dimanan de no haber sabido ser prudentes y energéticos durante un momento histórico, que no volverá jamás"* [1], adelantándose a lo que está sucediendo en la actualidad. Siendo evidente que nos encontramos ante una nueva era, la cual no podemos pasar por alto, debemos tomar medidas para ser prudentes y no caer en esos "males de pueblos e individuos" que dijera el Premio Nobel de Fisiología y Medicina. Resulta evidente que, como cualquier avance revolucionario, la inteligencia artificial tiene el mismo potencial para beneficiar a las personas que para lastimarlas.

Es importante reconocer que no siempre entendemos completamente el significado de la inteligencia artificial, tanto desde el punto de vista conceptual como del ontológico. Necesitamos entender qué es realmente la inteligencia artificial y tener una comprensión profunda y una definición precisa de la misma que nos permita establecer una base sólida.

Estamos constatando hoy en día cómo este debate sobre los dilemas éticos relacionados con la inteligencia artificial está cada vez más presente, sobre todo en lo que respecta al uso de herramientas de Big Data y los riesgos políticos asociados con su mal uso, que pueden llevar a la manipulación y distorsión de procesos sociales. En el ámbito del Derecho, la llegada de la tercera ola de la Inteligencia Artificial ha generado un cambio de paradigma que, de no ser manejado adecuadamente, podría plantear importantes problemas ético-jurídicos, especialmente en lo que respecta a la protección de los derechos de las personas más vulnerables (García San José et al., 2021). Hay que señalar que son muchos los derechos fundamentales que se ven afectados por la inteligencia artificial. Por ejemplo, la discriminación puede verse exacerbada por algoritmos sesgados, resaltando la importancia de la transparencia en el funcionamiento de la inteligencia artificial (García, 2023).

[1] Ramón y Cajal, S., *Charlas de café. Pensamientos, anécdotas y confidencias, 1920.*

No cabe duda de que ,en el contexto del Estado de Derecho, el uso de la inteligencia artificial debe estar sujeto al principio de legalidad, lo que implica cumplir con las normas jurídicas y las regulaciones establecidas. Nos enfrentamos a varios desafíos, siendo el principal la necesidad de regular su desarrollo, mediante disposiciones legales y no simplemente basándose en consideraciones éticas. Pero evidentemente no puede dejarse de lado la perspectiva no solo ética, sino filosófica, puesto que la IA, tal y como se caracteriza en este momento y teniendo en cuenta las pretensiones de algunos desarrolladores, no constituye una mera herramienta, sino un fenómeno con el potencial de revolucionar nuestra forma de vida. Históricamente hablando, los mayores descubrimientos de la especie humana han constituido siempre un arma de doble filo. El dominio del fuego nos proporcionó calor y seguridad, pero también permitió el refinamiento de los útiles de guerra. En un principio la invención de la maquina de vapor solo parecía beneficiosa para la humanidad, hasta que años mas tarde descubrimos su impacto en nuestro entorno natural. Podemos teorizar y elucubrar sobre los impactos próximos que puede tener la IA, pero no podemos ni imaginarnos las repercusiones que tendrán en la estructura de nuestra sociedad dentro de, por ejemplo, 50 años.

Si se tiene en cuenta el auténtico potencial de la IA, es sencillo ver cómo no estamos tratando ante un posible problema solo para el Estado de Derecho. El desarrollo de la IA nos acerca al debate sobre cómo podría el ser humano convivir con otros tipos de entes inteligentes y autónomos, y cómo la existencia de dichos entes afectaría a la concepción y caracterización que como especie hemos desarrollado sobre nosotros mismos.

CAPÍTULO II. FUNDAMENTOS TEÓRICOS

1. CONCEPTO DE INTELIGENCIA ARTIFICIAL

Si ya resulta complejo de por sí definir el concepto de la inteligencia, más arduo es conceptualizar la inteligencia artificial (IA). El término IA hace referencia a un tipo de máquinas y sistemas con ciertas capacidades, pero existe controversia sobre cuáles son sus rasgos definitorios. Algunos autores focalizan la esencia de la IA en su capacidad de aprendizaje continuo, otros en su manera de imitar los procesos cerebrales humanos. De forma general, la inteligencia artificial puede definirse como el conjunto de prácticas, principios y herramientas científicas que buscan emular, a través de una máquina, las capacidades cognitivas de los individuos, partiendo de datos aportados por el diseñador, el usuario o recolectados por el propio dispositivo desde fuentes digitales o sensoriales del entorno físico[2].

Los sistemas de IA, en su intento de replicar los mecanismos lógicos humanos, integran y encarnan una serie de características comunes[3]:

1. Adaptabilidad y aprendizaje: A diferencia de otras tecnologías que solo pueden funcionar en base a los contenidos de su programación original, las IAs aprenden a partir de los datos que se les aportan, adaptando su funcionamiento y volviéndose más sofisticadas y complejas con el paso del tiempo. Estas características hacen que algunos se pregunten si dicha capacidad de aprendizaje está limitada o si la IA puede continuar aprendiendo hasta volverse más inteligente que los propios humanos. Dentro de este denominado aprendizaje mecánico, existe el aprendizaje profundo, que describe la forma mediante la cual los sistemas de IA utilizan una red de capas superpuestas de neuronas artificiales que permiten el procesamiento simultáneo de grandes cantidades de datos, imitando la estructura y funcionamiento del cerebro humano.
2. Autonomía: Una de las características más intrínsecas al propio concepto de la IA es la autonomía de sus sistemas, pues estos están diseñados para funcionar sin requerir una constante intervención humana. Una vez diseñada y entrenada, la IA puede desarrollar sus tareas sin ayuda o supervisión humana, y si su configuración lo

[2] Elisa Gutiérrez García, *Inteligencia artificial y derechos fundamentales*, 2024, p. 23.
[3] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems, 25, 1097-1105

permite, ella misma podrá buscar datos que continuar procesando para perpetuar su aprendizaje y funcionamiento indefinidamente.

3. Racionalidad: El funcionamiento de la IA se basa en su capacidad de aplicar procesos lógicos a la información que se le proporciona, y formular conclusiones razonables a partir de estos.
4. Percepción e Identificación: Las herramientas de IA son capaces de detectar los estímulos en su entorno e identificar patrones en conjuntos de datos. La proliferación de sistemas de IA de reconocimiento biométrico es una de las principales preocupaciones en el debate sobre el uso de la IA.

El punto de partida histórico más citado para el surgimiento de la Inteligencia Artificial es la Conferencia de Dartmouth de 1956 (Sanabria Navarro et al., 2023), liderada por Minsky, Mc Carthy y Shannon. Sin embargo, es crucial reconocer fundamentos previos que sentaron las bases para esta disciplina. En 1948, Norbert Wiener, estableció los principios de la cibernética (Brunet Icart & Morell Blanch, 2001). Él mismo se basó en dicho concepto para describir una innovadora idea sobre el control y la comunicación en tanto máquinas como organismos vivos, al afirmar que, "*Hemos optado por denominar cibernética a todo lo relacionado con el control y la teoría de la comunicación, tanto en máquinas como en animales".* Si bien el término como tal no había sido acuñado con anterioridad, la noción de este ya fue planteada por los filósofos presocráticos. Prueba de ello la encontramos en la obra de Aristóteles, *"Física",* en la cual alude a las ideas de Anaximandro sobre el principio fundamental del universo, indicándonos que el infinito no tiene un origen definido, sino que parece ser el principio de todos los demás seres, abarcándolos y gobernándolos. Además, considera que lo infinito posee características divinas al ser inmortal e indestructible. Estas afirmaciones indican que la sustancia primaria que abarca y gobierna todo es una materia infinita y primordial, inseparable del movimiento planteando así como el infinito ejerce control sobre todos los seres.

Volviendo al siglo XX, en 1949, Hebb introdujo el paradigma conexionista al presentar su trabajo sobre los mecanismos de aprendizaje. A su vez, en 1950, Turing propuso la famosa prueba de Turing, un criterio para evaluar la inteligencia de una máquina basado en su capacidad para emular el comportamiento humano. Esta prueba plantea situar a un humano y una máquina en habitaciones separadas, mientras un observador les hace preguntas a través de una puerta. Si el observador no puede distinguir quién es quién después de un tiempo determinado, se concluye que la máquina exhibe inteligencia. Estos hitos históricos subrayan

el progreso continuo hacia el desarrollo de la Inteligencia Artificial (Parra Sepúlveda & Concha Machuca, 2021).

La inteligencia artificial es un proyecto de la ciencia iniciado hace 70 años pero que aloja en su seno una pulsión utópica que conecta con los orígenes utópicos y deterministas de la ciencia, concretamente con el *Nuovum Organum* de Francis Bacon y con el desarrollo que alrededor de esta idea plantea Thomas Hobbes en *De Homine,* donde se introduce la idea de que el conocimiento es poder y que, al ser un poder, tiene una capacidad de acción sobre el mundo que puede transformar este sobre la propia capacidad de poder que genera, y que en sí mismo tiene una energía transformadora. Dicha energía ha de ser necesariamente virtuosa porque se supone que el conocimiento tiene que contribuir al progreso y a un desarrollo positivo del mundo.

En 2023, representantes de distintos estados y organizaciones internacionales como la UE se reunieron en el Reino Unido para debatir sobre las implicaciones éticas del desarrollo de la IA, produciendo la conocida como Declaración de Bletchley. Dicho documento recoge una serie de principios y postulados encaminados a propiciar una gestión responsable y colaborativa del fenómeno de la IA.

Los países asistentes a la congregación acordaron que la colaboración y el diálogo internacional son vitales para asegurar una regulación de la IA ética y segura, aceptando que es un fenómeno destinado para cambiar para siempre las dinámicas humanas y cuya regulación, por tanto , debe ser fruto de un consenso de la mayor parte de la sociedad, siendo necesario crear unos estándares básicos y compartidos sobre el propósito, la moral y los límites de la IA.

La colaboración internacional iría encaminada a asentar unas bases éticas de la IA basadas en principios como la no discriminación y la igualdad de acceso, pero, fundamentalmente, dicha base ética debería estar orientada desde el antropocentrismo, es decir, transcurra como transcurra el desarrollo de la IA, esta debe permanecer siempre como una herramienta en beneficio de la sociedad y el ser humano, no como algo que amenace su propia existencia. Para ello es necesario encontrar el punto de equilibrio entre los beneficios aportados por la IA y la protección de los derechos de los ciudadanos, para lo que es absolutamente necesario cierto grado de supervisión e involucración humana en los procesos de IA. Es imprescindible crear un sistema de responsabilidad respecto a los actos de la IA que permita dirigirse a los desarrolladores e implementadores de esta.

En definitiva, la Declaración de Bletchley pone en relevancia que el desarrollo de la IA es un asunto de tal importancia que trasciende las fronteras, ofreciendo a la humanidad en su totalidad oportunidades al igual que amenazadas, y debiendo por tanto ser tratado como un asunto de vital importancia, en el que la colaboración internacional para preservar el beneficio humano es imprescindible.

José María Lassalle, en su libro *Civilización artificial*[4], distingue dos modelos de IA. Por un lado, la Nihilista, surgida en Estados Unidos y China, y diseñada para el poder sin propósito, ya sea hacia el control estatal o corporativo. En China, está vinculada al paradigma de Estado-plataforma, mientras que en Estados Unidos, se basa en el mercado digital y el egoísmo. Por otro lado , el modelo Humanista que encuentra su base en Europa, y está centrado desde una perspectiva teórica en lo humano y ético, pero que según Lassalle aún se enfrenta a desafíos para incorporar dimensiones éticas en la regulación.

2. ESTADO DE DERECHO: DEFINICIÓN Y PRINCIPIOS FUNDAMENTALES

Se dice que no hay democracia sin Estado de Derecho.

El Estado de Derecho es una de las creaciones intelectuales más asombrosas de la humanidad. En primer lugar, supone que el mismo poder que emite las normas consiente en ser autolimitado por estas normas de manera que éstas se convierten en leyes abstractas que regulan tanto el comportamiento del propio estado como el de los ciudadanos, es decir, el Estado se autolimita. El Estado de Derecho es por tanto un principio de gobernanza que establece que tanto la sociedad civil como el propio Estado están limitados por una ley abstracta que establece los límites de su actuación.

En segundo lugar, el Estado de Derecho establece la interdicción de la arbitrariedad de los poderes públicos, los cuales deben estar sujetos en su actuación a esa legislación abstracta. Dicho corpus legislativo habrá de ser público y conocido, de forma que todos tengan una previsibilidad de las consecuencias de sus actos, y deberá aplicarse en base a los principios de igualdad y no discriminación, así como de proporcionalidad.

El Estado de Derecho se configura en base a la separación de poderes, de forma que cada una de sus tres ramas podrá actuar de forma independiente dentro de sus limitadas capacidades, evitando de esta forma una acumulación excesiva de poder en ninguna institución estatal.

[4] Lassalle, J. M. (2018). Civilización Artificial. Editorial Taurus.

Los poderes públicos, en sus respectivas áreas de actuación, deberán velar por el mantenimiento y aplicación de los derechos fundamentales, asegurando la dignidad y la libertad de los ciudadanos.

2.1 España como un Estado Democrático de Derecho

La calificación del Estado español como democrático de Derecho aparece consagrada en la Constitución Española de 1978, cuyo art. 1.1 dispone que *"España se constituye en un Estado social y democrático de Derecho, que propugna como valores superiores de su ordenamiento jurídico la libertad, la justicia, la igualdad y el pluralismo político"*[5] . Esta primera referencia actúa como piedra angular en la arquitectura constitucional, plasmando la preocupación del legislador constituyente por construir unas sólidas y estables instituciones representativas[6].

A su vez, en el art. 9.1 se establece que, "*Los ciudadanos y los poderes públicos están sujetos a la Constitución y al resto del ordenamiento jurídico"*[7] y se complementa con la salvaguarda del art. 103 que expresa cómo la estructura del Estado y sus entidades públicas, respaldadas por la Constitución, garantizan su protección[8]. Por último, en el art. 117 se subraya que la administración de justicia emana del pueblo, lo que implica la existencia de tribunales independientes sujetos exclusivamente a la ley[9].

3. INTERSECCIÓN ENTRE INTELIGENCIA ARTIFICIAL Y ESTADO DE DERECHO

La democracia digital, definida como el fortalecimiento de la democracia tradicional mediante tecnologías de la información y comunicación, puede mejorar la participación ciudadana y la calidad de los procesos democráticos. Sin embargo, es importante reconocer que no garantiza la participación política por sí sola y debe complementarse con medidas fuera del ámbito digital.

[5] Art. 1 CE

[6] ContrerasM. (1986). Iniciativa legislativa popular y estado democrático de derecho (una aproximación a su regulación jurídica en España). *Revista De Las Cortes Generales*, (8), 67-94. https://doi.org/10.33426/rcg/1986/8/445

[7] Art. 9 CE

[8] Art. 103 CE

[9] Art. 117 CE

Puede aplicarse en procesos donde la ciudadanía participa en la elaboración de normativas, tanto para recopilar datos e información como para evaluar la calidad de los procesos participativos. Sin embargo, es crucial abordar la desigual capacidad de intervención ciudadana que puede surgir del uso de la IA.

Por ello, es fundamental que los sistemas de IA utilizados por las Administraciones Públicas estén sujetos a control jurisdiccional para garantizar su legalidad y evitar posibles abusos, y que las decisiones administrativas basadas en la IA sean susceptibles de ser revisadas por los tribunales. Además, es primordial garantizar que el desarrollo y uso de la IA respeten los derechos civiles, democráticos y sociales de los ciudadanos, lo que requiere un enfoque equilibrado que proteja tanto la eficiencia como los derechos fundamentales.

Debemos aspirar a un entorno digital seguro, que garantice a protección de la privacidad y los datos personales al utilizar herramientas de democracia. En el ámbito electoral, se ha demostrado que, aunque en España y en otros países se utilizan herramientas electrónicas para agilizar los procesos electorales y el voto, estos métodos aún no son significativamente más efectivos que los métodos tradicionales en términos de participación ciudadana. La IA plantea desafíos adicionales en la democracia digital, especialmente en lo que respecta al control de plataformas tecnológicas dominantes y la manipulación de datos personales para influir en las decisiones de los votantes. Por lo tanto, se requieren estrictas obligaciones de transparencia y medidas efectivas de cumplimiento para mitigar estos riesgos.

En relación con el carácter social del Estado y su búsqueda de la igualdad real, es crucial evitar que los sesgos en la era digital perjudiquen a personas y grupos sociales. Se destaca la importancia de establecer regulaciones concretas, como las establecidas en la Ley 15/2022, que exigen a las administraciones públicas considerar criterios de minimización de sesgos, transparencia y rendición de cuentas en el uso de algoritmos en la toma de decisiones.

Además, el acceso efectivo de la ciudadanía al mundo digital depende de la educación y la financiación proporcionada por los poderes públicos. Se insta a los Estados miembros a integrar la alfabetización digital en la educación básica y el aprendizaje continuo, con énfasis en la necesidad de un enfoque interdisciplinario que abarque tanto aspectos técnicos como humanísticos y sociales.

La dimensión social de los derechos, junto con el principio democrático, requiere una consideración amplia de la eficacia de estos derechos, tanto verticalmente (frente a los poderes públicos) como horizontalmente (entre individuos dentro de la sociedad), teniendo en cuenta el impacto de la informática, la IA y las tecnologías relacionadas en las relaciones laborales y en el ámbito del consumo.

En el ámbito laboral, se enfatiza la necesidad de transparencia por parte de los empleadores en el uso de la IA, así como la protección de los derechos laborales de los trabajadores frente a decisiones basadas en algoritmos.

Además, en el ámbito del consumo, se requieren normas de protección específicas, incluidas la identificación de la interacción con dispositivos de IA y la disponibilidad de mecanismos de revisión humana. Es crucial prestar atención al sector de los seguros, donde si bien la IA puede ofrecer beneficios significativos, también plantea importantes riesgos, como la exclusión y la discriminación (Presno Linera, 2023).

El estudio del derecho y la tecnología de la información conlleva una contradicción inherente, ya que mientras la tecnología abarca nociones como la internacionalización y la globalización, el derecho, en su mayor parte, todavía está confinado a las fronteras nacionales. Trascendiendo en cierta medida esta contradicción a la noción del Estado de Derecho, que conlleva el ideal de que "el Estado de Derecho es bueno para todos", una actitud que aparentemente goza de apoyo internacional.

Hay que ser conscientes del peligro que puede suponer la tecnología que contiene elementos de inteligencia artificial , ya que conforme esta disciplina avanza y se consiguen herramientas más sofisticadas, estas herramientas se infiltran en nuestros sistemas de toma de decisiones, de forma que los sistemas de toma de decisiones digitales amenazan con desplazar a aquellos en los que intervienen los humanos, debido al afán de estos últimos por conseguir la mayor eficacia posible.

La mayoría de estos sistemas de toma de decisiones son "*cajas negras*" porque incorporan tecnología extremadamente compleja que está más allá de las capacidades cognitivas humanas y del derecho (Brożek et al., 2024). Es aquí donde las exigencias del Estado de Derecho, como la visión, la transparencia, la justicia y la aplicabilidad, son casi imposibles de lograr, lo que a su vez plantea preguntas sobre hasta qué punto el Estado de Derecho es un concepto viable en la sociedad tecnocrática.

La tecnología se considera frecuentemente como una "espada de doble filo" porque sus impactos en la sociedad pueden ser tanto positivos como negativos. Por ejemplo, aunque puede limitar la libertad de expresión, también puede promoverla.

El caso Wisconsin- Loomis, en Estados Unidos ilustra algunos de los peligros que presenta la aplicación de la IA. La Administración de Justicia adoptó el algoritmo COMPASS, diseñado para, en base a datos tanto generales como específicos del caso, determinar el riesgo de reincidencia del individuo en diferentes etapas del proceso penal. Esto daba siempre como resultado un porcentaje de reincidencia más elevado en las personas de raza negra. Se rechazó en apelación, señalando cuatro motivos a tener en consideración; primero, que se trata de un software privado y por tanto carece de transparencia, al no conocerse todas las exactitudes de su funcionamiento; segundo, el sistema COMPASS evalúa el riesgo de reincidencia en grupos sociales, no en individuos concretos; tercero, utiliza datos nacionales en lugar de datos específicos de Wisconsin; y cuarto, diversos estudios demuestran que los algoritmos de sentencia tienden a clasificar desproporcionadamente a los delincuentes minoritarios como de mayor riesgo.

Un aspecto esencial del Estado de Derecho es que las leyes deben ser accesibles para que las personas puedan cumplirlas y entender lo que se espera de ellas, siendo la previsibilidad crucial. No basta con regirse por leyes, sino que estas no producirán sus efectos hasta ser debidamente publicitadas, y solo en el caso de que la redacción de su contenido permita la comprensión por parte de los ciudadanos sobre los que opera. Por eso, conceptos como la publicación y la inteligibilidad, defendidos por académicos como Fuller, son fundamentales para el Estado de derecho. La IA, con su falta de accesibilidad, puede socavar estos principios. El lenguaje natural en el que se ha basado el Estado de Derecho para ser comprendido, y el lenguaje de la IA o lenguaje algorítmico son totalmente distintos, y no puede exigirse a la ciudadanía (al menos en este preciso momento) una comprensión igual del segundo que del primero. Evidentemente, al igual que no se espera del ciudadano medio que conozca con exactitud la totalidad del corpus legislativo, no puede esperarse una absoluta comprensión de los complejos algoritmos que utiliza la IA, pero sí es exigible, cuando esta se utilice en la administración de justicia, suficiente claridad y transparencia en los mecanismos que conducen a los resultados aportados por la IA.

El artículo 63 del Reglamento General de Protección de Datos (RGDP) que amplía el artículo 22 sobre IA, otorga a los interesados el derecho a conocer y recibir información sobre la

lógica detrás del procesamiento de datos en decisiones automatizadas, proporcionando una posible explicación de la tecnología. Sin embargo, este derecho se ve limitado cuando los secretos comerciales y los derechos de propiedad intelectual prevalecen sobre la transparencia. El RGPD es un ejemplo de cómo la ley tradicional intenta equilibrar los derechos de propiedad intelectual con los derechos de privacidad en relación con el uso de la IA.

Esto se muestra igualmente en el artículo 15(1)(h) del RGPD, que da a los interesados el derecho a información sobre la lógica del procesamiento automatizado y sus consecuencias.

Sin embargo, como venimos referenciando, la naturaleza compleja y opaca de la IA, dificulta este equilibrio, ya que la transparencia, necesaria para proteger adecuadamente contra sus posibles daños, choca con los derechos de propiedad intelectual que pueden contribuir a crear una "caja negra" tecnológica.

Aunque los derechos de propiedad intelectual y los secretos comerciales están protegidos para fomentar la creatividad y ofrecer incentivos económicos, se crea un conflicto con los principios de transparencia y publicidad, en un caso más de choque entre la protección de los derechos privados e individuales con la protección general de la ciudadanía.

Para conseguir corregir el desequilibrio creado por la IA, se pueden emplear diversos mecanismos, como puede ser recurrir a terceros de confianza que pueden ayudar a asegurar que los algoritmos empleados se desarrollen y utilicen conforme a los valores del Estado de Derecho. En este sentido, en el Reino Unido, la Comisión de Derecho sobre el Uso de Algoritmos en el Sistema de Justicia recomendó la creación de un Registro Nacional de Sistemas Algorítmicos para verificar los algoritmos utilizados en el sistema de justicia penal.

Según Krygier, el Estado de Derecho tiene como objetivo principal hacer que la ley gobierne, para frenar el potencial de abuso de poder por parte de aquellos que lo utilizan de manera arbitraria. Él afirma que hay muchas formas de ejercer el poder y que las maneras arbitrarias deben ser rechazadas.

Siguiendo este razonamiento, la Comisión de Venecia de "prevención del abuso (mal uso) del poder" sostiene que hay una correlación entre, por un lado, definir el Estado de Derecho a través del prisma del poder y, por otro lado, la noción de reciprocidad. Para que la reciprocidad sea efectiva, se requiere un cierto equilibrio en la relación de poder entre quienes gobiernan y los gobernados. Sin embargo, se advierte que con la transferencia de la

gobernanza a la tecnología, como se observó en el caso Loomis, se puede caer en el monopolio en términos de acceso a la tecnología.

No se puede obviar que son los gobiernos los que ~~t~~ienen los recursos para producir o comprar la tecnología que se utiliza para tomar decisiones sobre los ciudadanos. Este desequilibrio creciente supone que se está despojando de poder a los gobernados en favor de los que gobiernan, pudiendo acabar los ciudadanos inmersos en un sistema en el que herramientas que no terminan de comprender deciden sobre su inocencia, sus condenas y la protección de sus derechos.

A su vez, la Comisión de Venecia advierte de que esta monopolización del poder sobre la tecnología en manos de quienes gobiernan aumenta el riesgo de que el poder del ejecutivo se vuelva ilimitado, contrariando al Estado de Derecho y dando lugar a posibles abusos de poder o comportamientos injustos y antidemocráticos.

La Comisión apunta finalmente otro aspecto importante a tener en consideración respecto a este ansiado equilibrio de poder Los productores de la tecnología de IA son actores privados, por lo que la ecuanimidad debe lograrse entre tres entidades: quienes gobiernan, los gobernados y las corporaciones privadas que desarrollan la tecnología para la mediación, con el riesgo de que haya situaciones en las que los actores privados ejerzan poderes que tradicionalmente han sido ejercidos por los Estados. Como ya se ha referido, esto supone un caso más de tratar de encontrar el equilibrio entre la libertad de empresa y la protección de la propiedad intelectual que mantiene el mercado competitivo, y la protección de los derechos individuales de los ciudadanos.

Otro punto de contacto entre el uso de la IA y el Estado de Derecho radica en la separación de poderes que establece este último. Si se tiene en cuenta la creciente implantación de las herramientas de IA en todos los sectores de la vida moderna, puede imaginarse un futuro cercano donde opere en las instituciones que representan los tres distintos poderes configurados en nuestro actual Estado. Si a esto se suma la preocupación sobre la posibilidad de que la IA devenga autoconsciente y autónoma, podríamos enfrentarnos a un Estado en el que los poderes solo quedan separados desde la perspectiva de los seres humanos, mientras que la IA se configuraría como un ente omnipresente en todas las ramas de aplicación de los distintos poderes.

4. RIESGOS DE LA INTELIGENCIA ARTIFICIAL

El lanzamiento en el año 2023 de Chat GPT-4, una herramienta dotada de la capacidad de sintetizar y producir texto, ha suscitado un considerable interés acerca de cómo la inteligencia artificial generativa remodelará el ámbito laboral, el acceso a los servicios gubernamentales y la interacción en las plataformas de Internet en general (*Global*, 2023).

A pesar de que, como toda innovación tecnológica, la inteligencia artificial generativa puede generar oportunidades, la ausencia de una regulación adecuada y eficaz podría agravar los riesgos para los derechos humanos en diversos campos, como por ejemplo en el acceso a la protección social, a la educación y el empleo, respecto a los derechos laborales, la privacidad y la seguridad digital, entre otros.

Entre los peligros potenciales, se encuentra el riesgo de acentuar el racismo y a aumentar el discurso de odio en Internet. Las tecnologías de inteligencia artificial existentes, entre otras, ya han exacerbado la desigualdad y perjudicado a comunidades marginales en áreas como el acceso a los servicios estatales, la actuación policial, la seguridad y la migración (*Digitally Divided*, s. f.).

Por ejemplo, en Serbia, la implementación del nuevo sistema semiautomatizado de protección social, respaldado por el Banco Mundial, provocó la pérdida potencial del acceso a una asistencia social para posiblemente miles de personas y afectó de manera desproporcionada a la población romaní y a las personas con discapacidad.

Otro ejemplo lo tenemos con el empleo de la tecnología de reconocimiento facial por parte de Israel en los Territorios Palestinos Ocupados, que ha aumentado las restricciones a la libertad de circulación y ha contribuido a mantener el sistema de apartheid.

En el continente americano en 2023, concretamente en Nueva York, el Departamento de Policía admitió haber utilizado esta tecnología para vigilar las protestas del movimiento Black Lives Matter en la ciudad, lo que incrementó las presiones sobre el Ayuntamiento de Nueva York para que las prohibiera (*Silicon shadows: Venture capital, human rights, and the lack of due diligence - Amnesty International*, s. f.).

No son pocos los que claman que el miedo hacia la IA es desmedido e irracional, y que tratar de limitar y condicionar su funcionamiento supone tratar de frenar el progreso. Muchos opinan que los escenarios que se plantean para debatir sobre la correcta regulación de la IA son irreales y distópicos. Sin embargo, basta con poner el foco en el uso que el gobierno

chino está haciendo de los sistemas de IA para ser consciente de lo necesario que es imponer principios éticos sólidos al desarrollo e implantación de la IA.

El gobierno chino utiliza los sistemas de IA para reforzar su autoritario control sobre la población de diversas formas. La IA juega un importante papel en mantener la censura en la red, así como en impulsar la propaganda gubernamental. Sistemas de IA escanean y monitorizan la actividad cibernética, identificando mensajes subversivos y contrarios al gobierno, denunciando a sus autores y eliminando o enmascarando el contenido, a la par que otras herramientas crean y difunden propaganda en cantidades masivas con el fin de manipular la opinión pública. Aunque el gobierno chino lleva décadas censurando a sus detractores y realizando campañas propagandísticas, es indudable que las herramientas de IA facilitan en sumo la labor de manipulación de la información y supresión de la libertad de expresión.

La combinación de la extensa red de videovigilancia existente en China con las herramientas de IA de reconocimiento facial y análisis biométrico resulta en un gobierno omnisciente comparable con el Gran Hermano Orwelliano. No solo permite la detección inmediata de los comportamientos calificados como indeseados y la identificación inmediata de sus autores, sino que constituye un fuerte medio de represión ciudadana, sabiéndose la población vigilada y monitorizada perpetuamente, y desprovistos de cualquier derecho a la intimidad o la privacidad. Ni siquiera es necesario que los ciudadanos sean descubiertos mientras cometen los delitos, pues las autoridades chinas ya están implantando sistemas predictivos (completamente sesgados y discriminatorios) para la determinación *ex ante* de la inocencia o culpabilidad de los ciudadanos. La población uigur, históricamente estigmatizada y perseguida, se ve particularmente afectada por los datos inexactos y discriminatorios que utilizan estas herramientas para elaborar sus perfiles forenses (Westphal & Wang, 2023).

En definitiva, el caso de China muestra a la perfección la necesidad de someter el desarrollo e implementación de los sistemas de IA a principios y límites éticos sólidos y consensuados, que sean fruto de una reflexión profunda y holística, que aseguren la protección de los derechos y libertades de los ciudadanos.

Sin embargo, los riesgos que plantea la IA van mucho más allá que estos ejemplos de casos concretos. Muchos de los axiomas y postulados en torno a los que, como especie, hemos basado nuestros sistemas éticos, legales y, en general, en los que hemos basado nuestra existencia, han derivado históricamente de nuestra condición de "creaciones". Dicha condición, común al el resto de los seres vivos con los que compartimos planeta, nos ha mantenido como especie sujetos a ciertos límites. Sin embargo, conforme las tecnologías avanzan y nos acercan a una posible realidad donde la IA puede hasta desarrollar consciencia propia, nos aproximamos a un momento determinante en nuestra historia como especie en el que podemos alcanzar al final la condición de "creadores", alterando permanentemente nuestra propia concepción de nosotros mismos.

Realmente, la posibilidad de que la IA desarrolle conciencia de su propia existencia plantea cuestiones de suma importancia. ¿Cómo debería el ser humano tratar a un producto de su propia creación? Si la humanidad como especie ha tardado más de 12,000 años en determinar cómo deben tratarse las distintas razas y nacionalidades de nuestro planeta, y aun así no hallamos consenso en cómo debemos tratar al resto de especies de seres vivos, ¿conseguiremos decidir en cómo tratar a un ente producto de nuestra propia acción?

Si la IA avanza hasta el punto en el que esta pueda ser consciente de sus actos y consecuente al respecto, ¿deberíamos proteger jurídicamente su autonomía? ¿Podría hablarse de voluntad en sus actos tal y como se plantea hoy en día el concepto en nuestro Ordenamiento Jurídico? ¿Dónde acabaría la responsabilidad de aquel que maneja la herramienta y dónde empezaría la de esta?

En su obra "Yo, Robot", Isaac Asimov estableció las primeras normas que deberían regir el comportamiento de máquinas conscientes y autónomas, pero nada se dijo respecto a los derechos que ostentarían . Y es más, si se tiene en cuenta el empeño que algunas personas tienen en conseguir que la IA no solo imite el cerebro humano, sino que supere sus capacidades, ¿cómo estableceremos el límite entre la protección de la IA y la seguridad de la humanidad en su conjunto? ¿Cómo limitaríamos el rango de actuación de un ente autónomo y mas inteligente que nosotros mismos?

Es importante tener en cuenta en el debate sobre el desarrollo de la IA el concepto de la dignidad. Ciertamente, la mayoría de las consideraciones y reglas éticas que rigen nuestra sociedad hoy en día derivan del concepto de la dignidad humana. En base a la dignidad humana, defendemos los derechos y obligaciones que el hombre debe ostentar, establecemos el límite entre lo permitido y lo prohibido, lo justo y lo injusto, lo decente y lo inhumano.

En base a la dignidad humana, Kant estableció (y como especie, estuvimos de acuerdo) que el ser humano es un fin en sí mismo, nunca un medio, derivando de esta conceptualización gran parte de nuestro sistema de derechos fundamentales y universales. Por lo tanto, y teniendo en cuenta que multitud de autores asocian la idea de dignidad a la capacidad de autoconsciencia y reflexión ética, ¿podríamos (y deberíamos) admitir la existencia de dignidad en algo que no nació como un fin, sino como un medio? ¿Y qué derechos y deberes derivarían de esa "dignidad artificial"?

Como puede verse, el desarrollo de la IA no solo afecta al Estado de Derecho en cuanto a leyes de protección de consumidores o de responsabilidad de las IA se refiere, sino que puede transformar por completo todas las consideraciones éticas y jurídicas en torno a las que hemos basado no solo nuestra forma de vida, sino nuestra concepción propia como especie.

CAPÍTULO III. DESAFÍOS Y PREOCUPACIONES ÉTICAS

1. SESGOS Y DISCRIMINACIÓN ALGORÍTMICA

El sesgo tiene lugar "cuando los datos disponibles no son representativos de la población o fenómeno en estudio y cuando los datos no incluyen variables que capturen adecuadamente el fenómeno que queremos predecir y que los datos incluyen contenido producido por humanos que puede contener sesgo contra grupos de personas".

Las IAs utilizan datos tanto para su entrenamiento y aprendizaje como para la aportación de sus resultados. Desde el momento en que manipulan datos, éstos traen consigo sesgos, por lo que éstos son una parte inherente de la ciencia de datos y de las tecnologías de IA.

Hay que señalar que la mayoría de los conjuntos de datos tienen sesgos, intencionados o no, y siempre es asumible algún grado de sesgo, que inevitablemente da como resultado salidas inexactas por los modelos entrenados con dichos datos. Simplemente la elección de un conjunto de datos sobre otro puede reflejar un sesgo específico. La Agencia de Derechos Fundamentales de la Unión Europea (FRA) emitió en 2022 un informe sobre los sesgos en los algoritmos, describiendo cómo estos pueden darse en varios momentos del ciclo vital de la IA, pudiendo darse en la propia recolección de datos, en el posterior aprendizaje en base a estos, y evidentemente en la elaboración de resultados. Por ello, el informe insta a articular mecanismos de control y vigilancia que aseguren no solo la inexistencia de prejuicios en los resultados ofrecidos por la IA, sino que la información con la que se alimenta a la IA sea de calidad, no discriminatoria y mínimamente sesgada.

En este punto es importante distinguir entre sesgo y discriminación. Esta última es un concepto legal que se refiere al "trato prejuicioso de una persona basado en su pertenencia a un grupo o categoría específica", y abarca atributos como raza, religión, nacionalidad, género, orientación sexual, discapacidad, estado civil, características genéticas, idioma y edad. Por lo tanto, aunque puede admitirse la presencia de un nivel mínimo de sesgo, en el momento en el que de la aplicación de este resulta un trato especialmente diferente hacia colectivos o grupos de personas concretos, estaríamos hablando de un sistema discriminatorio y por lo tanto inadmisible según los postulados del Estado de Derecho.

La Comisión de Venecia considera la igualdad ante la ley y la no discriminación como elementos esenciales del Estado de Derecho, abarcando motivos como raza, color, sexo, idioma, religión, opinión política, origen nacional o social, asociación con una minoría nacional, propiedad, nacimiento u otro estatus. Al igual que considera la discriminación

opuesta al Estado de derecho, demandando no solo la no discriminación sino también "igualdad en la ley" e "igualdad ante la ley"[10]. Esta distinción es relevante para sistemas de decisión con IA, donde puede ser difícil identificar discriminación específica debido a la complejidad matemática de los modelos.

El problema que subyace con el sesgo y la discriminación en el contexto de los datos es el "enmascaramiento" que ocurre cuando dos características están correlacionadas y se usa una característica trivial para indicar un característica sensible, considerándose un modelo discriminatorio por ejemplo si dos personas con las mismas características relevantes para un proceso de toma de decisiones reciben diferentes resultados debido a un atributo sensible. Es importante señalar que, aunque el sesgo en los datos puede llevar a efectos discriminatorios, esto no siempre es el caso.

Además, otro inconveniente a considerar es que las reglas matemáticas de estos modelos no han sido necesariamente expuestas a los procedimientos legislativos tradicionales, sino que son creadas por corporaciones privadas, lo que dificulta la tarea de asegurar que se ajustan a los parámetros de igualdad de trato y no discriminación.

2. RESPONSABILIDAD LEGAL DE LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL

Es innegable que la introducción de la inteligencia artificial en el mundo del Derecho conlleva aspectos positivos, pero también presenta importantes desafíos.

Su naturaleza singular, su complejidad y la imposibilidad de conocer, de antemano, el comportamiento de la máquina ha generado la necesidad de establecer una normativa específica, ya que las reglas tradicionales no se adaptan adecuadamente a esta nueva realidad, como, por ejemplo se ha demostrado que ocurre, en asuntos de responsabilidad por actos de terceros y atribución de culpa.

De ahí que surja la propuesta de un Reglamento del Parlamento Europeo y del Consejo, de 21 de abril de 2021, que establece normas armonizadas sobre inteligencia artificial (Ley de Inteligencia Artificial), aprobada preliminarmente en diciembre de 2023. Aunque no aborda el tema de la responsabilidad civil, a través de requisitos, condiciones, demandas,

[10] European Commission for Democracy Through Law (Venice Commission), *Report on the Rule of Law*, CDL-AD (2011)003rev

prohibiciones y, en particular, mediante la clasificación de los sistemas, incorpora un marco para garantizar un uso responsable de esta tecnología.

En este contexto, se propone una definición única para estos sistemas, que sea lo suficientemente amplia y adaptable a los cambios futuros. El artículo 3.1 de la propuesta en cuestión (enmienda 165) subraya que se trata de "*un sistema basado en máquinas diseñado para funcionar con diversos niveles de autonomía y capaz, para objetivos explícitos o implícitos, de generar información de salida –como predicciones, recomendaciones o decisiones– que influya en entornos reales o virtuales*". Algunos autores (Fernández, 2024) incluyen dentro de los subcampos de la inteligencia artificial el aprendizaje automático, el aprendizaje profundo, el procesamiento del lenguaje natural y la visión por computador.

Además, se establecen ciertas prohibiciones según las distintas aplicaciones de la inteligencia artificial, así como una serie de requisitos para los sistemas de inteligencia artificial considerados de alto riesgo, pudiendo resumir los aspectos más relevantes de dichas prohibiciones y requisitos en los siguientes puntos:

- Prohibiciones sobre sistemas de IA que interactúan con personas físicas: Opera la regla fundamental de que dichos sistemas deberán informar a los usuarios de que están tratando con herramientas de IA y no con personas físicas.
- Prohibiciones sobre sistemas de reconocimiento de emociones: En aras de proteger la intimidad y privacidad de los usuarios, estos sistemas se someterán a fuertes restricciones, como la prohibición de su uso en lugar de trabajo.
- Prohibiciones sobre sistemas de categorización biométrica: Además de quedar limitado su uso a ciertas áreas, deberán garantizarse estrictos mecanismos de seguridad que aseguren la protección de los datos ante ataques externos, debido al carácter sumamente personal de estos.
- Prohibiciones sobre sistemas de IA para generar o manipular contenido: Se configuran estrictas condiciones para el uso de IA en la creación o manipulación de contenido audiovisual, con el objetivo de disminuir la desinformación, la difamación y la manipulación social.
- Requisitos para sistemas de IA de alto riesgo: Estos sistemas se sujetarán a condiciones y restricciones que garanticen su transparencia, equidad y no discriminación. Por ejemplo, deberá asegurarse que los sistemas de IA utilizados para el diagnóstico medico sean lo suficientemente rigurosos y exactos en sus resultados.

- Normas de seguimiento y supervisión del mercado: Deberán implantarse mecanismos de supervisión y seguimiento del uso de las IA, de forma que se asegure el cumplimiento del resto de condiciones y requisitos.

Por lo tanto, las disposiciones especiales relacionadas con el derecho de daños deben considerar la mencionada "norma horizontal" para desarrollar un sistema de responsabilidad civil coherente con los principios de la Unión Europea. Dado que el funcionamiento del mercado requiere el constante intercambio de productos y servicios entre los países miembros (y con otros fuera de la Unión) y que esta actividad puede implicar riesgos, la legislación debe actuar en dos momentos diferentes: previamente, mediante normas sobre seguridad de los productos, para eliminar, reducir o minimizar los posibles daños causados; y retrospectivamente, cuando el producto haya causado perjuicios para compensar a la víctima.

Estas tres facetas no deben ser consideradas como compartimentos estancos, sino que están intrínsecamente interconectadas. En este sentido, las regulaciones sobre la seguridad de los productos, al igual que las normativas de alcance horizontal, contribuyen a mitigar el riesgo de ocurrencia de perjuicios, aunque no lo erradiquen por completo. De igual manera, las disposiciones sobre seguridad se presentan como un complemento de las normas de responsabilidad civil, siendo, en esencia, dos aspectos complementarios de una misma materia, ya que aunque no resuelvan por sí solas las controversias asociadas con la responsabilidad, son altamente relevantes para su determinación, pues son consideradas al establecer su existencia. En resumen, ambas operan en momentos diferentes y se refuerzan mutua y necesariamente.

A parte de las disposiciones nacionales y de aquellas promulgadas en el ámbito de la Unión Europea y que aplican en sectores específicos, destaca la Directiva del Consejo de 25 de julio de 1985, que trata la armonización de las normas legales, reglamentarias y administrativas de los Estados miembros en cuanto a la responsabilidad por daños causados por productos defectuosos. No obstante, la inteligencia artificial presenta desafíos para su adecuada inserción en esta Directiva, dado que los productos dotados de inteligencia artificial poseen un funcionamiento que difiere de la definición contenida en la misma. Precisamente debido a esta circunstancia, se ha procedido a revisar y proponer una nueva Directiva sobre responsabilidad civil por productos defectuosos que permita su aplicación en el ámbito de la IA, así como una Directiva relacionada con la adaptación de las normativas de responsabilidad civil extracontractual a la inteligencia artificial.

Una de las cuestiones esenciales que subyacen a estas propuestas es la dificultad probatoria que podría sufrir la víctima al tratar de demostrar los elementos alegados en la demanda, incluyendo la culpa del operador de un sistema inteligente, su funcionamiento defectuoso, el nexo causal y el daño ocasionado. Además, otro desafío que adquiere relevancia en este ámbito es la multiplicidad de actores involucrados, quienes podrían ser considerados potenciales responsables.

Desde esta óptica, ambos instrumentos buscan abordar estas discusiones. En el caso de la propuesta de Directiva sobre responsabilidad civil por productos defectuosos, la cuestión de la culpa no constituye un verdadero obstáculo, dado que incorpora un régimen objetivo de responsabilidad. La cuestión central radica en la determinación de la defectuosidad del producto y la causalidad. En este sentido, mientras el artículo 6 de la propuesta de Directiva establece los criterios para identificar un producto defectuoso, los artículo 8 y 9, mediante la asignación de la carga de la prueba al demandado y los requisitos de las normativas de seguridad, establecen ciertas presunciones tanto sobre la defectuosidad como sobre la relación causal. De esta manera, en términos generales y con ciertas salvedades, se puede inferir que si el demandante no cumple con la mencionada obligación de presentar pruebas, se presumirá que el producto es defectuoso y que existe una relación directa entre éste y el daño sufrido.

En lo que respecta a los sujetos responsables, la propuesta de Directiva sobre responsabilidad civil por productos defectuosos introduce diversas reglas de asignación de responsabilidad para garantizar que la víctima siempre pueda identificar un responsable dentro de la Unión Europea. Estas reglas se aplican de manera preferente y subsidiaria, lo que significa que el primer responsable de indemnizar el daño es el fabricante; en su ausencia (si está establecido fuera de la Unión), el importador y el representante autorizado; y en su ausencia, el proveedor de servicios de tramitación de pedidos a distancia; y así sucesivamente. A esto se suma el mantenimiento de la responsabilidad solidaria que ya estaba contemplada en la versión original de la Directiva.

En el caso de la propuesta de Directiva relativa a la adaptación de las normativas de responsabilidad civil extracontractual a la inteligencia artificial, es importante destacar que también incluye disposiciones destinadas a mitigar estos efectos negativos. Asimismo, en este contexto surge una problemática específica, dado que la aplicación de la responsabilidad subjetiva implica que el demandante también deba demostrar la culpa (Reyes López, 2023).

Como destaca Astray Chacón (Fernández, 2024), la adaptación normativa, que se impone como necesaria, se proyecta sobre tres instrumentos de armonización: la propuesta de Reglamento 2021, la propuesta de Directiva sobre responsabilidad civil extracontractual de la IA y la propuesta de revisión de la Directiva sobre responsabilidad por los daños causados por productos defectuosos.

Con base al análisis llevado a cabo por la Comisión Europea, se concluye que la mayoría de las normativas nacionales aplican un régimen subjetivo de responsabilidad, si bien se identifican posibles dificultades para probar el daño, lo que motiva la búsqueda de formas de aligerar la carga probatoria. En este contexto, la Unión Europea reconoce que el denominado efecto de "caja negra" puede dificultar o encarecer excesivamente para las víctimas la determinación de la persona responsable y la demostración de los requisitos para una demanda de responsabilidad civil viable. Además, la adaptación de las normativas por parte de los órganos judiciales podría generar disparidades, lo que resultaría en una clara inseguridad jurídica. Por consiguiente, se propone la intervención de la Unión Europea para evitar la fragmentación y la disminución de la inversión económica en este sector.

En cuanto a la propuesta de Directiva, algunos autores sugieren que, de ser aprobada, podría integrarse fácilmente en los sistemas de responsabilidad civil existentes, ya que refleja un enfoque que no modifica conceptos fundamentales como "culpa" o "daño," los cuales varían considerablemente entre los Estados miembros. Compartiendo esta opinión, se considera pertinente incluir una definición sobre los daños indemnizables, dada la gran diversidad existente entre los Estados miembros y las particularidades del sector en cuestión, tanto en términos de cuantía de los perjuicios como de la amplia variedad de daños.

Es relevante destacar que, al igual que en la propuesta de Directiva sobre productos defectuosos, el artículo número tres de la propuesta de Directiva para la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial establece el régimen de exhibición de pruebas. En este sentido, dicho artículo dispone que los Estados miembros deben garantizar que los órganos judiciales tengan la facultad de ordenar, a solicitud del demandado y previa presentación de hechos y pruebas suficientes, la exhibición de las pruebas pertinentes de las que disponga. Sin embargo, esta medida debe limitarse a lo estrictamente necesario y realizarse de manera proporcional, considerando la posible presencia de información confidencial, secretos comerciales y otros derechos de propiedad industrial o intelectual. En este contexto, el demandante puede solicitar medidas para preservar la confidencialidad, las cuales pueden ser adoptadas de oficio por el juez.

Además, en atención al tipo de responsabilidad subjetiva aplicable, se establece una presunción de culpabilidad en caso de incumplimiento. Sin embargo, esta regla general presenta matices según el tipo de Sistema y la consideración del demandado. Específicamente, en el caso de sistemas de alto riesgo, de acuerdo con las disposiciones sobre gestión de riesgos y sus resultados contemplados en la propuesta de Reglamento del Parlamento Europeo y del Consejo de 21 de abril, se presume la existencia de culpa en los casos descritos en el artículo 4.2 de la propuesta de Directiva para la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial.

En cualquier caso, cuando se enfrente a un Sistema de alto riesgo, no se supondrá automáticamente la culpa si el demandante tiene la condición de usuario, entendido como "implementador" según la terminología propuesta por el Reglamento del Parlamento Europeo y del Consejo de 21 de abril de 2021, es decir, si su uso se limita a una actividad personal no profesional, y el demandado demuestra que no cumplió con sus obligaciones de utilizar o supervisor el Sistema de acuerdo con las instrucciones, o si lo expuso a datos de entrada irrelevantes para su finalidad.

CAPÍTULO IV. MARCO REGULATORIO Y POLÍTICAS PÚBLICAS

1. LEGISLACIÓN NACIONAL SOBRE INTELIGENCIA ARTIFICIAL Y DERECHOS FUNDAMENTALES

En el contexto del Estado democrático, es fundamental reconocer la importancia del pluralismo, el respeto a la minoría y la participación ciudadana en las distintas funciones del Estado, como establecen los principios constitucionales del artículo 9.2 de la Constitución Española.

En esta dirección, en España, la Estrategia Nacional de Inteligencia Artificial (ENIA) propone establecer un marco ético y normativo para fortalecer la protección de los derechos individuales y colectivos, promoviendo la inclusión y el bienestar social.

El documento conocido como la Declaración de Toronto, suscrita el 16 de mayo de 2018, emitido por organizaciones como *Human Rights Watch, Access Now* y Amnistía Internacional, marca un hito en la historia al buscar aplicar los estándares internacionales de derechos humanos al desarrollo y uso de sistemas de aprendizaje automático, también llamado inteligencia artificial.

En su esencia, la declaración insta a gobiernos y empresas a asegurar que las aplicaciones de aprendizaje automático e inteligencia artificial se rijan por los principios de igualdad y no discriminación. Establece directrices de derechos humanos que tanto el sector público como el privado deben seguir para garantizar una aplicación justa y equitativa de los algoritmos, así como proporcionar vías de reparación significativas para aquellos cuyos derechos hayan sido violados. Además, reafirma el derecho a un recurso efectivo y a responsabilizar a los responsables de dichas violaciones, enfatizando la importancia de la Responsabilidad en cuanto al uso de la IA se refiere. También hace un llamamiento a los gobiernos para asegurar el debido proceso en el uso de aprendizaje automático en el ámbito público, ejercer cautela en la implementación de sistemas de IA en el sistema penal, establecer mecanismos claros de rendición de cuentas en el desarrollo e implementación de aplicaciones de IA, y definir claramente quienes son los responsables legales de las decisiones tomadas mediante el uso de tales sistemas.

Además, reconoce que la inteligencia artificial es una herramienta estratégica con beneficios significativos para ciudadanos, empresas y la sociedad en general, siempre y cuando se

oriente hacia el ser humano, sea ética y sostenible, y respete los derechos y valores fundamentales[11].

A su vez, destaca cómo el aprendizaje automático y la IA afectan a una amplia gama de derechos humanos, como la privacidad, la libertad de expresión y el derecho a la vida. Uno de los principales riesgos del aprendizaje automático es amplificar la discriminación y los sesgos existentes, lo que puede tener consecuencias perjudiciales para grupos marginados y vulnerables. Ejemplos como la discriminación algorítmica en la concesión de hipotecas ilustran este problema (Grigore, 2022).

La Alta Comisionada de las Naciones Unidas para los Derechos Humanos, Michelle Bachelet, ha llamado a establecer una moratoria sobre la venta y uso de sistemas de IA que amenacen los derechos humanos hasta que se implemente salvaguardas adecuadas. La transparencia, la equidad de género en el desarrollo de la IA y la evaluación constante de su impacto en los derechos humanos son fundamentales para mitigar riesgos y garantizar su uso ético y responsable.

La legislación canadiense, por ejemplo, establece normas para la toma de decisiones automatizadas, exigiendo transparencia, pruebas para detectar sesgos y mecanismos para la intervención humana y la impugnación de decisiones administrativas automatizadas.

El Parlamento Europeo ha enfatizado la importancia de abordar los riesgos asociados con la toma de decisiones basada en la IA y ha pedido un marco normativo que garantice una gobernanza efectiva y protección de los derechos fundamentales.

1.1. Directrices éticas de la UE para una IA confiable

En 2019, la Comisión Europea emitió una serie de directrices y recomendaciones para orientar el uso de la IA. Estas disposiciones no vinculantes adoptan un enfoque antropocentrista para describir una serie de principios éticos que aseguren un uso confiable y adecuado de la IA. Puede hablarse de siete principios clave:

- Intervención y supervisión humanas: Es necesaria la supervisión humana para asegurar que la IA no socaba nuestra propia autonomía ni genera efectos indeseados en nuestra persona. En el uso de la IA deben garantizarse mecanismos de control

[11] Santiago Camilo Carretero Sánchez, *La responsabilidad del Estado en sus nuevos frentes: sanitario, alimentario, energético y de inteligencia artificial*, 2023.

basados en la adaptabilidad, la exactitud y la aplicabilidad de los sistemas de IA, manteniendo una constante participación humana basada en el control y supervisión.

- Solidez y seguridad técnicas: Los sistemas de IA deben ser lo suficientemente sofisticados y desarrollados como para asegurar la integración de mecanismos de control y seguridad que aseguren tanto la veracidad de sus resultados como la protección ante ataques externos. Deberá asegurarse la mayor fiabilidad y reproducibilidad posible de sus datos, así como la reversibilidad de las consecuencias no deseadas derivadas de su funcionamiento.
- Privacidad y gestión de datos: Para que los ciudadanos se sientan suficientemente confiados como para proporcionar a la IA sus datos personales, es fundamental garantizar una adecuada protección y gestión de estos. Las personas deben mantener en todo momento el control sobre sus datos, y ser aseguradas de que no se usarán para fines desfavorables o ilícitos. No solo deberá garantizarse que los datos están seguros ante ataques externos o manipulaciones, sino que deberán eliminarse los sesgos e inexactitudes de los conjuntos de datos empleados para entrenar a la IA, de forma que no ofrezca posteriormente resultados erróneos.
- Transparencia: El funcionamiento interno de los sistemas de IA deberá ser comunicado a sus usuarios, tratando en la medida de lo posible de dar luz sobre los procesos internos que conducen a los resultados aportados por la IA. Es fundamental que la actuación de la IA sea guiada por la trazabilidad, de forma que queden registradas no solo las decisiones tomadas por estas herramientas, sino los procesos que han seguido para llegar a ellas.
- Diversidad, no discriminación y equidad: La IA debe evitar sesgos, tanto en los datos en base a los que se educa y entrena, como en los resultados que aporta. Los sistemas de IA deberán tener en cuenta todo el espectro de capacidades y aptitudes humanas, de forma que su diseño sea lo más accesible e igualitario posible.
- Bienestar social y medioambiental: El impacto de las herramientas de IA no debe medirse desde una perspectiva individual, sino social, valorando si su empleo beneficia o perjudica a la sociedad en su conjunto.
- Rendición de cuentas: Los sistemas de IA deberán integrar mecanismos que permitan la auditoría de estos, en especial en sus aplicaciones que puedan afectar a los derechos fundamentales. Así mismo, deberá garantizarse la responsabilidad y la rendición de cuentas tanto de los propios sistemas de IA como de los resultados ofrecidos por estos.

1.2.Carta de Derechos Digitales

La Carta de Derechos Digitales, presentada en 2021, constituye una directriz no vinculante encaminada a describir las situaciones de conflicto dadas entre el actual desarrollo tecnológico y la protección de los derechos individuales, para ofrecer así un marco ético y legislativo que permita consensuar la implantación de nuevas tecnologías como la IA en nuestra sociedad con la preservación de aquellos derechos, valores y principios derivados de la dignidad humana que sustentan el Estado de Derecho. Tal y como expresa la carta, su fin es el de ofrecer a los poderes políticos una contextualización de la situación actual y una serie de predicciones sobre el posible desarrollo futuro de las tecnologías que invite a un proceso de reflexión que permita, a su vez, aprovechar todo el potencial que ofrecen estas herramientas sin sacrificar los principios éticos que orientan nuestro corpus legislativo.

Admitiendo que nos desenvolvemos en un panorama rápidamente cambiante y en constante evolución, la carta no busca establecer criterios exactos para las tecnologías en actual desarrollo, sino crear un marco ético que subraye y proteja las dimensiones de la dignidad humana que puedan verse afectadas en el entorno digital.

La Carta de Derechos Digitales no busca crear nuevos derechos fundamentales, sino describir la relación de los ya existentes con el nuevo contexto tecnológico. Para ello, la carta recoge una serie de derechos proyectables en diversas áreas como los derechos laborales o la investigación científica, y en su sección vigesimoquinta, establece una serie de derechos ante la IA orientados desde la protección de la persona y su dignidad y la búsqueda del bien común. Consagra el derecho de los ciudadanos a una supervisión e intervención humanas en los procesos que empleen herramientas de IA, y a impugnar las decisiones tomadas por esta. La mayoría de los derechos y principios promulgados en la carta se reflejan en el marco ético que ésta crea para el uso de la IA, poniendo en relevancia el deber de utilizar sistemas accesibles y no discriminatorios que se basen en la transparencia, la trazabilidad, la supervisión humana y la gobernanza. Estos principios y derechos deberán ser especialmente garantizados cuando se aplique la IA en la Administración de Justicia.

En conclusión, la Carta de Derechos Digitales constituye un recordatorio directo de los principios y valores éticos que orientan la legislación española, el cual insta a los poderes públicos a dirigir la actividad legislativa en lo concerniente al desarrollo tecnológico desde el puro antropocentrismo, recordando que las nuevas herramientas tecnológicas deben servir y proteger al hombre y su dignidad.

1.3. El Real Decreto 729/2023, publicado en el BOE el 2 de septiembre de 2023, establece el Estatuto de la Agencia Española de Supervisión de Inteligencia Artificial (AESIA), que entra en vigor el 3 de septiembre de 2023.

Este estatuto aprueba la creación de la AESIA, el primer organismo de su tipo en Europa, en anticipación a la implementación del Reglamento Europeo de Inteligencia Artificial (Real Decreto 729/2023, de 22 de agosto, por el que se aprueba el Estatuto de la Agencia Española de Supervisión de Inteligencia Artificial, 2023).

El Real Decreto 729/2023 define la estructura organizativa de la AESIA, que incluye una Presidencia encargada de la dirección y representación de la agencia, un Consejo Rector como órgano colegiado para la toma de decisiones estratégicas, una Comisión Técnica de expertos para asesoramiento técnico y científico, y una Secretaría General responsable de la gestión administrativa y financiera.

Las funciones y competencias de la AESIA son amplias y abarcan varios aspectos críticos. La agencia tiene la responsabilidad de supervisar y evaluar el desarrollo y uso de la IA en diversos sectores, asegurando que estos se alineen con los principios éticos y legales vigentes que conforman nuestro Estado de Derecho. Además, debe evaluar el impacto de la IA en la sociedad y en los derechos fundamentales de los ciudadanos. En el ámbito regulatorio, la AESIA tiene la tarea de proponer y desarrollar normativas específicas para la implementación y uso de la IAE, garantizando que estas promuevan la seguridad, la transparencia y el respeto a los derechos humanos.

Otro rol esencial de la AESIA es brindar asesoramiento técnico a entidades públicas y privadas, promoviendo buenas prácticas y el cumplimiento de las regulaciones establecidas. La agencia también tiene un fuerte componente de colaboración internacional, fomentando la cooperación con organismos globales para el intercambio de conocimientos y la armonización de estándares en IA, y participando activamente en foros internacionales dedicados a la regulación de la IA. Como ya se ha descrito, la problemática asociada a la implantación de la IA en distintas áreas de la sociedad es un asunto de importancia global, que no debe relegarse a la mera ordenación nacional, al ser un fenómeno que afecta a la raza humana en su totalidad.

Además de su función reguladora, la AESIA tiene un papel importante en la promoción de la innovación, impulsando la investigación y el desarrollo en el campo de la IA, apoyando las iniciativas y proyectos que promuevan el avance tecnológico y contribuyan al crecimiento económico sostenible. Este enfoque busca asegurar que España no solo cumpla con los

estándares éticos y legales, sino que también se convierta en un líder en el desarrollo y uso responsable de la IA, contribuyendo a un desarrollo tecnológico enfocado desde la preservación y mejoramiento de la calidad de vida del hombre.

1.4. El Real Decreto 817/2023, publicado en el BOE el 9 de noviembre de 2023, establece un entorno controlado de pruebas para evaluar el cumplimiento de la propuesta de Reglamento del Parlamento Europeo y del Consejo sobre normas armonizadas en inteligencia artificial, que entra en vigor el 10 de noviembre de 2023.

Esta normativa pone en marcha el primer entorno controlado de pruebas para verificar cómo implementar los requisitos aplicables a los sistemas de Inteligencia Artificial de alto riesgo según la propuesta de Reglamento Europeo. Esta iniciativa se integra en la estrategia española de transformación digital, conocida como Agenda España Digital 2026, dentro del Plan de Recuperación y, específicamente, en la Estrategia Nacional de Inteligencia Artificial (Real Decreto 817/2023, de 8 de noviembre, que establece un entorno controlado de pruebas para el ensayo del cumplimiento de la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial, 2023). El objetivo de la Propuesta de Reglamento del Parlamento Europeo y del Consejo que establece normas armonizadas sobre Inteligencia Artificial (Acta de Inteligencia Artificial) y modifica ciertos actos legislativos de la Unión, es abordar las deficiencias de las reglas existentes, así como establecer un sistema de control efectivo para los sistemas de IA. Esta propuesta reglamentaria establece un enfoque basado en el riesgo y establece cuatro niveles de riesgo:

- Sistemas cuyo riesgo es tan inaceptablemente alto que están prohibidos (Título II)
- Sistemas que generan alto riesgo (Título III)
- Sistemas a los que, aunque no se consideren de alto riesgo, se aplican una serie de requisitos de transparencia (Título IV)
- Los sistemas restantes (Título IX)

La mayor parte de la propuesta de Acta de Inteligencia Artificial de la UE se centra en la regulación de los sistemas de alto riesgo. Dentro de los requisitos que deben cumplir estos sistemas, este documento se centra en los contenidos del artículo 10, que se refieren a los datos utilizados en el “entrenamiento, validación y prueba” de los sistemas de alto riesgo (Soriano Arnanz, 2023).

Se reconoce que la discriminación causada o mediada por el uso de sistemas automatizados ha sido identificada por la academia y las instituciones como uno de los principales riesgos derivados del creciente uso de algoritmos en muchas áreas de la vida económica y social (Directorate-General for Justice and Consumers (European Commission) et al., 2021).

Hasta la fecha no se ha podido dar una respuesta concluyente sobre los parámetros que deberían introducirse en un sistema automatizado para garantizar el respeto de los derechos a la igualdad y la no discriminación. Una de las principales razones por las cuales es extremadamente difícil establecer criterios fijos con los cuales todos los sistemas automatizados deben cumplir para considerarse no discriminatorios es la abstracción que caracteriza a las normas legales. En este sentido, dado que la interpretación y aplicación del marco normativo para la protección de los derechos a la igualdad y la no discriminación se realiza caso por caso, existe una variación significativa en los criterios aplicables dependiendo del contexto, el tipo de decisión tomada y las personas afectadas, entre otros elementos (Wachter et al., 2020). Aunque puede entrenarse a la IA con datos suficientes para tratar de tener todas las consideraciones particulares de cada caso en cuenta, aquí sale a relucir el debate sobre si la IA puede aplicar en sus procesos mecanismos propios del pensamiento humano como la empatía y la proporcionalidad, elementales para una apropiada administración de justicia.

A pesar de la dificultad, ya existen algunas reglas aplicables al uso de sistemas automatizados, como el artículo 22 del Reglamento General de Protección de Datos, que prohíbe en general las decisiones basadas únicamente en el procesamiento automatizado de datos.

1.5. El Reglamento Europeo de Inteligencia artificial (AIA, Artificial Intelligence Act)

El 21 de abril de 2024, el Consejo de la Unión Europea dio luz verde a la anticipada Ley de Inteligencia Artificial. Esta normativa supervisa la introducción en el mercado, la implementación y la utilización de sistemas de inteligencia artificial en la Unión Europea. Su propósito central es promover el crecimiento y la utilización de la IA dentro de la UE, apostando por el desarrollo tecnológico mientras garantiza una alta protección de la salud, la seguridad y los derechos fundamentales de los ciudadanos.

La normativa adopta un enfoque fundamentado en el riesgo, teniendo en cuenta los posibles riesgos derivados del uso de sistemas de inteligencia artificial, y establece requisitos y

obligaciones para los participantes en la cadena de valor. Estas obligaciones no se restringen únicamente a los proveedores de sistemas de inteligencia artificial, sino que también se extienden a otros actores, incluyendo aquellos que emplean dichos sistemas con fines profesionales, referidos como "implementadores".

La Ley de Inteligencia Artificial se extiende a los proveedores que introduzcan en el mercado o pongan en servicio sistemas de IA, así como modelos de IA de propósito general, dentro de la Unión Europea, sin importar si están establecidos o localizados dentro de la UE o en un tercer país. También aplica a los implementadores de IA que tienen su sede o están ubicados dentro de la UE; a los proveedores e implementadores de sistemas de IA establecidos o localizados fuera de la UE cuando los resultados producidos por dichos sistemas se utilicen dentro de la UE; a los importadores y distribuidores de sistemas de IA; a los fabricantes de productos que comercializan o ponen en servicio sistemas de IA junto con sus productos y bajo su propio nombre o marca; a los representantes autorizados de proveedores que no tienen sede en la UE; y a las personas afectadas que se encuentran dentro de la UE. Como puede verse, el campo de aplicación de la normativa es extenso y exhaustivo.

Esta Ley no se aplica a ciertos sistemas, tales como aquellos utilizados únicamente para actividades de investigaciones y desarrollo científico. Asimismo, queda excluida de su ámbito de aplicación cualquier actividad de investigación, prueba o desarrollo relacionada con sistemas de IA o modelos de IA antes de su comercialización o puesta en servicio. Además, los individuos que utilizan sistemas de IA exclusivamente para actividades personales no profesionales no están sujetos a la Ley de IA.

El artículo 3 de la Ley de Inteligencia Artificial define a un sistema de IA como,

> *"un sistema basado en una máquina que está diseñado para funcionar con distintos niveles de autonomía y que puede mostrar capacidad de adaptación tras el despliegue, y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales".*

Dicha definición, es la ya establecida por la OCDE, la cual la Ley IA adopta para así poder garantizar un entendimiento común y global de aquellos términos esenciales. Esta definición, y así la Ley de IA, excluye los sistemas de software que tienen capacidades menores a las mencionadas previamente.

Siguiendo su enfoque basado en el nivel de riesgo, la Ley de IA prohíbe completamente varias prácticas de IA que representan riesgos inaceptables. En concreto, la Ley de la IA prohíbe:

- Técnicas subliminales, manipulativas o engañosas que alteren el comportamiento de personas o grupos, llevándolos a decisiones que normalmente no tomarían y que probablemente les causen daño significativo.

- Explorar las vulnerabilidades de personas o grupos debido a su edad, discapacidad o situación socioeconómica, para distorsionar su comportamiento de manera que probablemente les cause daño significativo.

- Sistemas de IA que predicen delitos basándose únicamente en el perfil o los rasgos de personalidad de una persona.

- Sistemas de IA que expanden bases de datos de reconocimiento facial mediante la recopilación no autorizada de imágenes fáciles de Internet o CCTV.

- Sistemas de IA que infieren emociones en lugares de trabajo y educativos, salvo por razones médicas o de seguridad.

- Sistemas de IA que categorizan a personas basándose en datos biométricos para inferir datos sensibles.

- Sistemas de identificación biométrica remota en tiempo real en espacios públicos para la aplicación de la ley, con ciertas excepciones y garantías.

La Ley de IA clasifica como de alto riesgo ciertos sistemas que representan un riesgo significativo de daño para la salud, la seguridad o los derechos fundamentales. Se diferencia entre dos grupos:

1. Sistemas vinculados a la legislación de armonización de la Unión sobre la seguridad de los productos enumerados en el Anexo I de la Ley de IA: el sistema de IA será de alto riesgo cuando sea un producto incluido en esta legislación de armonización, o cuando sea un componente de seguridad de estos productos; y siempre que, bajo esta legislación de armonización, el producto o componente deba someterse a una evaluación de conformidad por una tercera parte.
2. Sistemas listados en el Anexo III de la Ley de IA: aquellos sistemas que, debido al área en la que se utilizan y al uso específico que se les da, en principio representan un

alto riesgo. El Anexo III establece ocho áreas, cada una de las cuales identifica casos específicos de uso que se consideran de alto riesgo (en la medida en que su uso no esté prohibido).

Estas áreas se resumen a continuación, junto con algunos de los usos identificados en cada una, sin entrar en gran detalle:

- Biometría: esto incluye sistemas utilizados para la verificación biométrica remota; la categorización biométrica basada en la inferencia de características sensibles o protegidas; y el reconocimiento de emociones, excepto en los casos en que esté prohibido.
- Infraestructura crítica: sistemas utilizados como componentes de seguridad en la gestión y operación de infraestructura digital crítica, tráfico vial, o en el suministro de agua, gas, calefacción o electricidad.
- Educación y formación profesional: sistemas utilizados para determinar la admisión a instituciones educativas y de formación profesional; para evaluar resultados de aprendizaje; para evaluar el nivel de educación al que una persona podrá acceder; y para monitorear y detectar comportamientos prohibidos de los estudiantes durante los exámenes.
- Empleo, gestión de trabajadores y acceso al autoempleo: sistemas de IA utilizados para (i) el reclutamiento o selección de personas, la colocación de anuncios de empleo dirigidos, el análisis y filtrado de solicitudes de empleo, y la evaluación de candidatos; y (ii) la toma de decisiones que afectan las condiciones laborales y la promoción o determinación de relaciones contractuales laborales; la asignación de tareas basada en el comportamiento individual o las características personales; y la monitorización y evaluación del desempeño y comportamiento de los trabajadores.
- Servicios privados esenciales y servicios públicos esenciales y beneficios: se refiere a los sistemas empleados para determinar el acceso a beneficios de asistencia pública fundamental y servicios; evaluar la solvencia crediticia de individuos (excepto aquellos sistemas destinados a la detección de fraudes financieros); realizar la evaluación de riesgos y la fijación de precios en seguros de vida y salud; y establecer la prioridad en el despacho de servicios de respuesta de emergencia, como los proporcionados por la policía, bomberos y asistencia médica, así como los sistemas de triaje de pacientes en emergencias sanitarias.

- Aplicación de la ley: comprende diversos sistemas utilizados para la prevención e investigación de delitos, incluyendo ciertos sistemas predictivos y aquellos destinados a la elaboración de perfiles de personas físicas o a la evaluación de la fiabilidad de las pruebas.
- Gestión de la migración, asilo y control fronterizo: abarca sistemas diversos utilizados para evaluar riesgos de seguridad, riesgos asociados a la migración irregular y riesgos sanitarios; para examinar solicitudes de asilo; y para la identificación de personas físicas en el contexto de la migración, el asilo o el control fronterizo.
- Administración de justicia y procesos democráticos: se refiere a sistemas empleados para asistir a una autoridad judicial en la investigación e interpretación de hechos y normativa jurídica, así como para influir en el resultado de una elección o referéndum, o en el comportamiento electoral de las personas físicas.

Aun cuando un sistema esté referido en el Anexo III, no será considerado de alto riesgo si no representa un riesgo significativo de perjuicio para la salud, la seguridad o los derechos fundamentales de las personas físicas, incluyendo la ausencia de influencia material en el resultado de la toma de decisiones. Para alcanzar una determinación, la Ley de IA establece condiciones específicas, que la Comisión podrá modificar en fecha ulterior. Asimismo, se solicita a la Comisión que proporcione directrices que incluyan "una lista exhaustiva de ejemplos prácticos de casos de uso de sistemas de IA que sean de alto riesgo y no de alto riesgo". En cualquier circunstancia, los sistemas de IA mencionados en el Anexo III que realicen perfiles de personas físicas serán siempre considerados de alto riesgo.

Si bien la aprobación de este Reglamento se ha celebrado como un hito por iniciar el largo proceso de legislación sobre las aplicaciones de la IA, y ha establecido algunas premisas y límites valorables, resulta un trabajo legislativo que deja mucho que desear en cuanto al fondo mismo de la IA. En un momento en el que la UE podría haber sentado un precedente sobre la propia conceptualización de la IA, se ha limitado a mantener una visión optimista sobre sus implicaciones y efectos, estableciendo algunos límites y presupuestos pero sin indagar en la naturaleza de la IA y cómo una perspectiva errónea sobre esta puede resultar fatal para la humanidad en su totalidad. Por lo tanto, se considera que sigue siendo absolutamente necesaria e imperativa una regulación de la IA que introduzca como elemento clave y definitorio de esta su propósito de contribuir al desarrollo y bienestar humano. Solo una vez haya quedado establecido el fin último que debe orientar la configuración y la actividad de la IA será lícito preocuparnos por las concreciones de sus distintas aplicaciones.

2. ESTRATEGIAS DE REGULACIÓN Y GOBERNANZA DE LA INTELIGENCIA ARTIFICIAL EN EL ÁMBITO LEGAL

Nos encontramos actualmente en la edad del *Big data,* un momento en el que después de años permitiendo la extracción y manipulación de nuestros datos en la red, han surgido herramientas como la IA que pueden utilizar dichos datos para fines previamente inimaginables. Por ello, estamos experimentando un cambio de paradigma en cuanto a una nueva generación de normativas de protección de datos, destacando el Reglamento General de Protección de Datos (RGPD) de la Unión Europea del año 2016[12]. Estas regulaciones se han adoptado progresivamente en diversos países, destacándose por su carácter proactivo y preventivo, involucrando a los individuos en el cumplimiento normativo y adaptándolo a cada caso específico, en consonancia con los principios de la innovación responsable.

Estas medidas se materializan, en primer lugar, en el principio de responsabilidad proactiva, que requiere que el responsable del tratamiento no solo cumpla con la normativa de protección de datos, sino que también sea capaz de demostrar su cumplimiento.[13] A este principio se suma un enfoque centrado en la gestión del riesgo, que implica la evaluación del riesgo para determinar las medidas de seguridad adecuadas según el nivel de riesgo identificado[14]. Ambos principios confluyen en la obligación de incorporar la protección de datos desde la etapa de diseño y por defecto, lo que implica la aplicación de medidas adecuadas considerando el estado actual de la tecnología, los costes involucrados y la naturaleza del tratamiento de datos, así como los riesgos para los derechos y libertades de los individuos afectados[15].

Además de estas medidas, se incluyen otras acciones complementarias que promueven la responsabilidad activa, como la designación de un delegado de protección de datos y los mecanismos de autorregulación a través de códigos de conducta y certificaciones, lo que evidencia que la normativa de protección de datos se encuentra a la vanguardia de este nuevo

[12] Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (DOU 4.5.2016 L 119/1)
[13] Art. 5 RGPD
[14] Art. 35 RGPD
[15] Art. 25 RGPD

enfoque de gobernanza del riesgo en el ámbito digital, estableciendo estándares que abarcan todo este dominio.

Es importante destacar que este cambio en la protección de datos no es aislado, sino que se enmarca en un contexto más amplio de transformación digital que ha generado cambios significativos tanto en la normativa específica aplicable a las tecnologías de la información y la comunicación (TIC), como en la normativa general sobre seguridad de productos y servicios y protección del consumidor, entre otros aspectos. Este cambio de paradigma refleja un progresivo abandono del enfoque *laissez-faire* tradicional frente a los avances tecnológicos, impulsado por la creciente preocupación por los riesgos asociados a ciertas innovaciones tecnológicas con potencial disruptivo, como las plataformas de servicios digitales y la inteligencia artificial.

En el caso específico de las grandes plataformas y la inteligencia artificial, se evidencia este cambio de paradigma a través de nuevas iniciativas impulsadas por la Unión Europea, como las propuestas de Leyes de servicios y de mercados digitales, que introducen una regulación más estricta previa al desarrollo de estas actividades, que hasta ahora habían operado bajo un régimen jurídico mínimo como servicios de la sociedad de la información.

No obstante, el establecimiento de la gobernanza de los riesgos digitales no puede concebirse de forma independiente de otros modelos de gestión de riesgos a nivel global. Por el contrario, gran parte de sus principios, instituciones y estructuras están siendo adoptados en la gestión de los nuevos riesgos surgidos de la transformación digital. Esto es evidente en el caso del principio de precaución, presente en la gobernanza de los riesgos digitales y que adquiere especial relevancia, como puede apreciarse en la propuesta de reglamento de la Ley de Inteligencia Artificial de la Unión Europea, que, aunque no lo menciona explícitamente, incorpora importantes medidas de intervención preventiva, como prohibiciones y obligaciones. Por consiguiente, no se puede afirmar que en el ámbito de la innovación, el principio de precaución haya sido desplazado por el principio de innovación, aunque este último haya emergido recientemente y carezca, por el momento, de reconocimiento formal en el Derecho europeo o nacional, limitándose su referencia a documentos informales que lo definen como un "impulso a la regulación que fomente la innovación manteniendo los estándares de protección de derechos y valores reconocidos por la Unión"[16]. Su falta de

[16] European Political Strategy Centre (EPSC), «Towards an Innovation Principle Endorsed by Better Regulation», EPSC Strategic Notes nº 14, 30 de junio de 2014.

consagración normativa no impide que el principio de precaución no inspire e influya en la actividad legislativa.

En cualquier caso, precaución e innovación no son principios antagónicos sino complementarios, ya que se busca evitar interpretaciones extremas que puedan frenar o desatar el desarrollo tecnológico, promoviendo medidas flexibles que se adapten a cada situación para hacer frente a los riesgos. El quid de la cuestión es encontrar el punto de equilibrio que permita promover el progreso y la innovación sin sacrificar la seguridad ciudadana. En el proceso de configuración del modelo de gobernanza de los riesgos digitales, se van delineando nuevos principios, instituciones y mecanismos que encuentran su justificación y aplicación específica en este ámbito, como las evaluaciones de riesgos que surgieron en el contexto de la protección de datos y que se están extendiendo a otras dimensiones de la innovaciones tecnológicas en el ámbito digital, bajo la denominación de evaluaciones de impacto. Estas evaluaciones ahora se están ampliando y desarrollando en otras ámbitos de la digitalización, como los sistemas de inteligencia artificial, donde adquieren nuevas funcionalidades al considerar aspectos más allá de la privacidad, como la discriminación o la exclusión, lo que implica la incorporación de estándares sociotécnicos para evaluar estos riesgos.

CAPÍTULO V. CONCLUSIONES

Las implicaciones de la Inteligencia Artificial en el Estado de Derecho en España son de una magnitud considerable, y su comprensión y abordaje requieren una atención meticulosa y multidisciplinaria. A lo largo de este estudio, se han explorado diversos aspectos relacionados con esta intersección entre la IA y el Estado de Derecho, desde el marco legal y regulatorio hasta los desafíos éticos y sociales que plantea la materia. Para abordar estos desafíos de manera efectiva, es necesario un enfoque holístico y colaborativo que combine el conocimiento y la experiencia de diversos actores y disciplinas. Solo así se puede garantizar que la IA se utilice de manera ética, responsable y compatible con los principios y valores del Estado de Derecho en España y más allá.

En primer lugar, cabe destacar que es innegable que la IA está transformando de manera vertiginosa diversos aspectos de la sociedad y la economía, incluido el ámbito jurídico. El desarrollo de la IA tiene el potencial para alterar de forma radical conceptos tan sustanciales como la conciencia, la humanidad, la dignidad y la consciencia, por lo que es indudable que su implementación en distintas áreas de la sociedad tiene un impacto directo en el Estado de Derecho, existiendo la posibilidad futura de tener que reestructurar éste por completo para dar cabida a un nuevo tipo de entes conscientes y autónomos. En España, como en muchas jurisdicciones, el desarrollo y la implementación de tecnologías basadas en el uso de la IA plantean desafíos y oportunidades significativas para el Estado de Derecho. Por un lado, la inteligencia artificial puede optimizar la eficiencia y la accesibilidad de la justicia, agilizando los procedimientos judiciales. Sin embargo, también suscita preocupaciones relacionadas con la privacidad, la equidad y la transparencia, que deben ser abordadas con sumo cuidado para garantizar la salvaguardia de los derechos fundamentales de los ciudadanos, sin obviar problemas más transcendentales como el debate sobre qué estatus jurídico otorgar a herramientas con consciencia y pensamiento propio.

En segundo lugar, el marco legal y regulatorio actual en España ofrece ciertas protecciones y salvaguardias en relación con el uso de la IA en el ámbito jurídico, aunque es imprescindible adaptarlo y fortalecerlo para hacer frente a los desafíos emergentes. Es imperativo establecer normativas claras y específicas que regulen el desarrollo, la implementación y el uso de sistemas de IA en el sistema judicial. Dichas normativas deben fijar estándares éticos y técnicos rigurosos, así como mecanismos de supervisión y control, para garantizar que la IA sea utilizada de manera justa, transparente y responsable, y que ésta constituya en todo caso una herramienta destinada a beneficiar al ser humano.

En tercer lugar, se destaca la importancia de promover la alfabetización digital y legal en la sociedad para fomentar una comprensión informada y crítica de la IA y sus implicaciones en el Estado de Derecho. Esto implica no solo proporcionar educación y capacitación en materia de tecnología y derecho, sino también promover el diálogo público y la participación ciudadana en la formulación de políticas relacionadas con la IA. La creación de espacios de deliberación y colaboración entre expertos en tecnología, derecho, ética y sociedad es esencial para garantizar que las decisiones en torno a la IA sean inclusivas, equitativas y democráticas.

En cuarto lugar, se requiere una colaboración estrecha entre los diferentes actores involucrados en el desarrollo y la implementación de la IA en el ámbito jurídico, incluidos los poderes públicos, el sector privado, la academia y la sociedad civil. Solo a través de un enfoque colaborativo y multidisciplinario se pueden abordar de manera efectiva los desafíos complejos que plantea la IA en el Estado de Derecho. Esto incluye la promoción de la investigación interdisciplinaria, la cooperación internacional y el intercambio de mejores prácticas y lecciones aprendidas.

En quinto lugar, es crucial reconocer que la IA no constituye una solución universal y que su implementación no está exenta de riesgos y limitaciones. Si bien puede mejorar la eficiencia y la precisión en la toma de decisiones judiciales, también puede introducir sesgos y errores inherentes a los algoritmos y los datos en los que se basa. Por lo tanto, es necesario adoptar un enfoque crítico y reflexivo hacia la IA, evaluando cuidadosamente sus beneficios y riesgos en cada contexto específico y asegurando que se utilice de manera complementaria y subordinada a los principios y valores fundamentales del Estado de Derecho. El diseño y la implantación de los sistemas de IA deberá ajustarse siempre a principios como el de transparencia, no discriminación, supervisión humana y rendición de cuentas, para asegurar que los ciudadanos son conscientes en todo momento del funcionamiento de las herramientas

a las que se someten, así como para garantizar que no sufren un trato discriminatorio y velar por su derecho a revertir los efectos lesivos derivados del incorrecto funcionamiento o uso de la IA.

Para finalizar, puede decirse que actualmente vivimos un momento clave en la historia de la humanidad, en el que como especie tenemos que reflexionar sobre qué nos hace humanos y sobre qué aptitudes y facultades queremos desarrollar en las nuevas tecnologías, para determinar qué áreas o mecanismos deberíamos reservar al ser humano. Nunca se había visto a la especie humana tan cerca de encontrar entes con una capacidad similar, y estos no son propios de la naturaleza, sino que es el propio *homo sapiens* quien los crea y dirige, en lo que algunos califican como un afán por alcanzar la divinidad. Mientras otros solo piensan en cómo conseguir que una máquina imite lo mejor posible el pensamiento y la mente humana, algunos se cuestionan si de verdad queremos dar a luz a creaciones con nuestra capacidad de análisis y comprensión, pero sin nuestra capacidad empática, nuestra ética y en general, nuestra esencia de humanidad. Nos encontramos ante una encrucijada cuyo fin puede alterar para siempre no solo el Estado de Derecho o el ordenamiento jurídico, sino el propio concepto de humanidad, por lo que conviene actuar con cautela y comenzar la labor legislativa de regulación de la IA estableciendo los límites de desarrollo que esta debe alcanzar, y conceptualizando la IA de tal manera que constituya una herramienta en beneficio del hombre.

BIBLIOGRAFÍA

1. LEGISLACIÓN

Constitución Española (BOE 29 de diciembre de 1978).

Ley Fundamental de la República Federal de Alemania (BOF 23 de mayo de 1949).

Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (DOU 4.5.2016 L 119/1)

Reglamento (UE) 2021/0106 del Parlamento Europeo y del Consejo de 21 de abril de 2024 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.

2. RECURSOS DE INTERNET

European Commission for Democracy Through Law (Venice Commission), Report on the Rule of Law, CDL-AD (2011)003rev. Recuperado 4 de junio de 2024, de https://www.venice.coe.int/webforms/documents/default.aspx?pdffile=CDL-AD(2011)003rev-e

Brożek, B., Furman, M., Jakubiec, M., & Kucharzyk, B. (2024). The black box problem revisited. Real and imaginary challenges for automated legal decision making. *Artificial Intelligence and Law*, *32*(2), 427-440. https://doi.org/10.1007/s10506-023-09356-9

Brunet Icart, I., & Morell Blanch, A. (2001). Epistemología y cibernética. *Papers. Revista de Sociologia*, *65*, 31. https://doi.org/10.5565/rev/papers/v65n0.1705

Cervantes, C. C. V. (s. f.). *CVC. Santiago Ramón y Cajal. Charlas de café: Capítulo X*. Instituto Cervantes. Recuperado 2 de mayo de 2024, de https://cvc.cervantes.es/ciencia/cajal/cajal_charlas/capitulo10.htm

Digitally Divided: Technology, Inequality, and Human Rights. (s. f.). Amnesty International USA. Recuperado 1 de mayo de 2024, de https://www.amnestyusa.org/reports/digitally-divided-technology-inequality-and-human-rights/

Fernández, M. O. (2024). La «adaptación» del derecho de daños a la inteligencia artificial: La propuesta de Directiva sobre responsabilidad. *IDP. Revista de Internet, Derecho y Política*, *40*, Article 40. https://doi.org/10.7238/idp.v0i40.419696

García, C. P. (2023). Miguel Ángel Presno Linera, Derechos fundamentales e inteligencia artificial. *DERECHOS Y LIBERTADES: Revista de Filosofía del Derecho y derechos humanos*, *49*, Article 49. https://doi.org/10.20318/dyl.2023.7729

García San José, D., Llano Alonso, F., & Villegas Delgado, C. (2021). Derechos humanos, Estado de Derecho e Inteligencia Artificial en la era digital. *IUS ET SCIENTIA*, *2*(7), 5-6. https://doi.org/10.12795/IETSCIENTIA.2021.i02.01

Global: Las empresas deben actuar de inmediato para garantizar el desarrollo responsable de la inteligencia artificial. (2023, junio 15). Amnistía Internacional. https://www.amnesty.org/es/latest/news/2023/06/global-companies-must-act-now-to-ensure-responsible-development-of-artificial-intelligence/

Grigore, A. E. (2022). Derechos humanos e inteligencia artificial. *IUS ET SCIENTIA*, *8*(1), Article 1. https://doi.org/10.12795/IETSCIENTIA.2022.i01.10

Inteligencia artificial y derechos fundamentales. (2024). https://biblioteca.colex.es/libros/inteligencia-artificial-y-derechos-fundamentales-una-convivencia-era-digital-7668

La responsabilidad del Estado en sus nuevos frentes: Sanitario, alimentario, energético y de inteligencia artificial. (2023). https://biblioteca.colex.es/libros/responsabilidad-estado-nuevos-frentes-sanitario-alimentario-energetico-inteligencia-artificial-5041

Parra Sepúlveda, D., & Concha Machuca, R. (2021). Inteligencia artificial y derecho. Problemas, desafíos y oportunidades. *Vniversitas*, *70*. https://doi.org/10.11144/Javeriana.vj70.iadp

Presno Linera, M. Á. (2023). Derechos fundamentales e inteligencia artificial en el estado social, democrático y digital de derecho. *Anuario de la Red Eurolatinoamericana de Buen Gobierno y Buena Administración*, *3*, 10.

Reyes López, M. J. (2023). La protección al consumidor al hilo de las nuevas propuestas legislativas comunitarias. *Actualidad civil*, *7*, 11.

Sanabria Navarro, J. R., Silveira Pérez, Y., Pérez Bravo, D. D., & Cortina Núñez, M. de J. (2023). Incidencias de la inteligencia artificial en la educación contemporánea. *Comunicar: Revista Científica de Comunicación y Educación*, *77*, 97-107.

Silicon shadows: Venture capital, human rights, and the lack of due diligence—Amnesty International. (s. f.). Recuperado 1 de mayo de 2024, de

https://www.amnesty.org/en/documents/POL40/7109/2023/en/?utm_source=annual_report&utm_medium=pdf&utm_campaign=2021