



# APM 3.9.

Publicación digital. - Asociación Profesional de la Magistratura

MARÍA JOSÉ RIVAS VELASCO

ANDALUCÍA ORIENTAL

## USO ÉTICO DE INTELIGENCIA ARTIFICIAL EN JUSTICIA\*

### I.- ¿Inteligencia Artificial?

### II.- Riesgos derivados del funcionamiento de los sistemas de IA.

#### II.I.- Sesgo algorítmico, discriminación.

II.II.- Falta de transparencia del algoritmo, explicabilidad de la decisión.

### III.- Riesgos éticos del uso de los sistemas de IA en Justicia.

#### III.I.- Uso de IA en el proceso de toma de decisiones.

III.I.I.- Seguridad jurídica, imparcialidad, derecho al juez natural.

#### III.I.II.- Opacidad de la decisión. Fakes.

#### III.I.III.- Eficientismo. Perfilado de jueces.

#### III.II.- Ética de los sistemas IA en Justicia.

### Resumen

Inteligencia Artificial fue la palabra del año 2022 según la RAE, y en el año 2023 según Bill Gates ha comenzado la era de su dominio<sup>1</sup>. La utilización de la máquina por el hombre que supuso una revolución en nuestra cultura, en la actualidad ha dado paso a la creencia que la máquina lo acabará sustituyendo, aunque, por el momento, únicamente facilita y agiliza procesos que, humanamente son

---

Esta ponencia ha sido publicada en el centro de documentación del CGPJ en el marco de encuentros con la Comisión de Ética Judicial con magistrado/as

<sup>1</sup> Blog Bil Gates <https://www.gatesnotes.com/The-Age-of-AI-Has-Begun>

lentos, falibles y repetitivos. Esperamos confiadamente que la IA superará nuestras limitaciones para seguir avanzando hacia, lo que entendemos, será un mundo híbrido, donde se aúnen y confundan las capacidades humanas y tecnológicas. Sin embargo, sin saber realmente qué es la IA y cómo funciona, ¿estamos dispuestos a ceder ámbito de libertad y derechos fundamentales reconocidos tras siglos de lucha a un sistema tecnológico elaborado por humanos y cuyo funcionamiento interno nos es ajeno?. ¿Realmente somos conscientes que la cesión permanente de nuestros datos alimenta al sistema?. Si las dudas expuestas las trasladamos al ámbito judicial, debemos cuestionarnos si estamos dispuestos a aceptar que los sistemas de IA puedan llegar a sustituir la decisión judicial. Y si la respuesta es positiva, debiéramos plantearnos si ¿es ético que se sustituyan las decisiones humanas por resoluciones adoptadas por IA?, ¿supondría esto ceder ámbito de soberanía pública?. ¿Cabría efectuar un reproche ético si, en pos de una ansiada agilidad y eficiencia, se pierde ámbito de independencia judicial?.

### **Palabras clave**

IA, sesgo, ética, regulación, transparencia, explicabilidad, derechos fundamentales, perfilado, eficientismo.

### **I.- ¿Inteligencia Artificial?**

No pretende, ni podría, este trabajo ofrecer una definición de lo que es la IA, ni tampoco es el objeto del mismo. Fundamentalmente, como se verá, porque tampoco la hay actualmente. Pero es preciso efectuar una aproximación al concepto para poder apreciar los riesgos de su uso. Como se ha indicado, pese al recurrente uso del vocablo, no existe una definición unánimemente aceptada, ya que mientras que para unos tiene que ver con formas de vida artificial que pueden superar la inteligencia humana, para otros, cualquier tecnología de tratamientos de datos puede considerarse como tal (recomendaciones personalizadas de contenidos en Netflix, Spotify, etc). Tampoco hay una definición oficial consensuada (un chiste informático, la define como cosas guays que los ordenadores no pueden hacer), al integrar en el concepto sistemas tecnológicos que realizan desde tareas que se nos antojan de una gran dificultad y que para el programa no la entrañan (como jugar al ajedrez, ejemplo de ello son las conocidas partidas de Deep Blue contra Kasparov en 1997), como otras, que a los humanos nos resultan totalmente naturales, como agarrar objetos, y que sin embargo para su ejecución por una máquina precisa de una coordinación de muchos actos de ejecutados modo simultáneo (vista del objeto, decisión del movimiento, precisión del agarre, decisión de la fuerza a emplear en función del peso del objeto, etc), de modo que precisa un software de robótica que combine las redes neuronales de aprendizaje profundo con la optimización de la planificación del movimiento para calcular como agarrar y mover objetos con suavidad y rapidez.<sup>2</sup>

Para efectuar un aproximación al concepto, emplearemos dos palabras clave descriptivas: Autonomía - capacidad de ejecutar tareas

---

<sup>2</sup> <https://noticiasdelaciencia.com/art/40250/el-aprendizaje-profundo-ayuda-a-los-robots-a-agarrar-y-mover-objetos-con-facilidad>

en situaciones complejas sin la dirección constante del usuario-, y adaptabilidad - capacidad para mejorar la ejecución de las tareas aprendiendo de la experiencia-, ahora bien, la utilización de la palabra *inteligencia* puede dar lugar a equívocos, porque se asocia a la idea que el sistema puede ser capaz de realizar tareas como un humano, y no es así. Dentro del concepto, se utilizan - integran - otros, que se refieren a campos conectados e interrelacionados (informática, estadística, matemática computacional y análisis de datos) en la IA, pero que según el objetivo perseguido, son utilizados todos o algunos de ellos:

.-Aprendizaje automático, que no es más que un subcampo de la informática que se refiere a sistemas que van mejorando a medida que se usan o que acumulan datos.

.-Aprendizaje profundo, al igual que el anterior es un subcampo de la informática y que cuando añade el término profundidad, se refiere a la complejidad del modelo matemático utilizado. El aumento de la capacidad y velocidad de computación de los ordenadores modernos ha permitido incrementar esa complejidad a los investigadores hasta alcanzar niveles cuantitativa y cualitativamente diferentes a los anteriores.

.-Ciencia de datos, disciplina que abarca varios subtipos que incluyen el aprendizaje automático, la estadística, aspectos en los que se utiliza la informática en concreto los algoritmos, el almacenamiento de datos y el desarrollo de aplicaciones web.

.-Robótica, construcción y programación de robots que puedan operar en situaciones complejas en el mundo real, exigiendo la intervención combinada de todos los ámbitos de la inteligencia artificial (visión, reconocimiento del habla, procesamiento de lenguaje natural, recopilación de información, razonamiento para procesar instrucciones y predecir consecuencias de las diferentes decisiones de acción, modelado cognitivo y/o computación afectiva).

La RAE define "inteligencia artificial" como la "disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico".

Por tanto lo que pudiera parecer que integra un concepto unitario, no es tal, ya que, como indica la disposición que está preparando la UE para la regulación de los sistemas de la IA<sup>3</sup>, otorgando una definición que pretende ser lo más tecnológicamente neutra posible y resistir al paso del tiempo lo mejor posible, habida cuenta de la rápida evolución tecnológica y del mercado en relación con la IA, considera el Sistema de inteligencia artificial (sistema de IA) como: *un sistema basado en máquinas que está diseñado para funcionar con distintos niveles de autonomía y que puede, con objetivos explícitos o implícitos, generar resultados tales como*

---

<sup>3</sup>Propuesta de REGLAMENTO DEL PARLAMENTO EUROPEO Y DEL CONSEJO POR EL QUE SE ESTABLECEN NORMAS ARMONIZADAS EN MATERIA DE INTELIGENCIA ARTIFICIAL (LEY DE INTELIGENCIA ARTIFICIAL) Y SE MODIFICAN DETERMINADOS ACTOS LEGISLATIVOS DE LA UNIÓN

predicciones, recomendaciones o decisiones que influyen en entornos físicos o virtuales<sup>4</sup>.

Considera por tanto como sistemas de IA:

.-Software: asistentes virtuales, software de análisis de imágenes, motores de búsqueda, sistemas de reconocimiento de voz y rostro.

.-Inteligencia artificial integrada: robots, drones, vehículos autónomos, Internet de las Cosas.

.-Sistemas de IA de propósito general: sistema de IA que puede ser usado para y adaptado a un amplio espectro de aplicaciones para el que no fue diseñado intencional y específicamente.

Integra el concepto, tanto los productos o servicios que emplean tecnologías de inteligencia artificial, como sistemas de inteligencia artificial independiente. Sin embargo, como se ha indicado, el uso extendido del vocablo conlleva la confusión de considerar IA cualquier sistema tecnológico de utilización masiva de datos. La IA puede ser entendida tanto como un conjunto de tecnologías que permiten emular, a través de las máquinas y algoritmos, algún tipo de comportamiento, cómo aquellas que llevan a cabo tareas, resolver problemas o ayudan en la toma de decisiones.

Sin embargo, no es inteligencia, como entendemos el concepto con la RAE, esto es facultad de la mente que permite aprender, entender, razonar, tomar decisiones y formarse una determinada idea de la realidad, ya que ni los neurocientíficos tienen muy claro lo que es. Pero lo cierto es que en la actualidad, no existe la denominada IA general, entendida como máquina que puede ocuparse y resolver cualquier tarea intelectual<sup>5</sup>. Actualmente los investigadores de este campo, han abandonado el ideal de la inteligencia artificial general debido a que apenas ha habido progreso a este respecto durante los últimos cincuenta años, a diferencia de la IA estrecha - que se ocupa

---

<sup>4</sup> (6) *La noción de sistema de IA en el presente Reglamento debe definirse claramente y ajustarse estrechamente a la labor de las organizaciones internacionales que trabajan en el ámbito de la inteligencia artificial para garantizar la seguridad jurídica, la armonización y una amplia aceptación, proporcionando al mismo tiempo la flexibilidad necesaria para adaptarse a la rápida evolución tecnológica en este ámbito. Además, debe basarse en características clave de la inteligencia artificial, como sus capacidades de aprendizaje, razonamiento o modelización, para distinguirla de sistemas de software o enfoques de programación más simples. Los sistemas de IA están diseñados para funcionar con distintos niveles de autonomía, lo que significa que tienen al menos cierto grado de independencia de las acciones respecto a los controles humanos y de capacidad para funcionar sin intervención humana.*

(6 bis) *Los sistemas de IA suelen tener capacidades de aprendizaje automático que les permiten adaptarse y realizar nuevas tareas de forma autónoma. El aprendizaje automático se refiere al proceso computacional de optimización de los parámetros de un modelo a partir de datos, que es una construcción matemática que genera una salida basada en datos de entrada. Los enfoques de aprendizaje automático incluyen, por ejemplo, el aprendizaje supervisado, no supervisado y de refuerzo, utilizando una variedad de métodos, incluido el aprendizaje profundo con redes neuronales.*

(6 ter) *Los sistemas de IA pueden utilizarse como sistemas de software autónomos, integrados en un producto físico (embebidos), utilizados para servir a la funcionalidad de un producto físico sin estar integrado en él (no incorporado) o utilizarse como componente de IA de un sistema más amplio. Si este sistema más amplio no funcionara sin el componente de IA en cuestión, todo el sistema más amplio deberá considerarse un único sistema de IA con arreglo al presente Reglamento*

<sup>5</sup> <https://dobetter.esade.edu/es/inteligencia-artificial-futuro>.

de una exclusiva tarea - que avanza a pasos agigantados<sup>6</sup>. En realidad como indica algún autor, se busca una mimesis entre la IA y las (neuro) ciencias cognitivas, ya que a la par que el cerebro es imaginado como un dispositivo computacional, sin embargo se pretende crear nuevo hardware con base en el modelo del cerebro, representado por esta tecnología<sup>7</sup>.

## **II.- Riesgos derivados del funcionamiento de los sistemas de IA.**

Como cualquier sistema elaborado por humanos, es falible y es preciso conocer que puede presentar disfunciones para poder emitir un juicio crítico sobre la conveniencia de su uso en Justicia, donde se ha de procurar en todo caso el respeto a la equidad en los resultados que muestra, la transparencia en el proceso de su formación, la explicabilidad de la decisión adoptada y el respeto a la privacidad y dignidad de las personas.

Conocer que la IA es actualmente capaz de alterar o generar contenidos de todo tipo (denominada IA generativa que evoluciona a optimización o verticalización de los modelos<sup>8</sup>), de modificar los datos de la biometría de una persona o de la voz (sistemas face2face o lyrebird) de modo que resulte imposible distinguir los creados artificialmente de los reales, nos obliga a cuestionar en nuestra función jurisdiccional el principio de inmediación, en cuanto máximo exponente de la primacía que se otorga a nuestra percepción sensorial, a la que atendemos y de la que dependemos necesariamente para valorar el contenido de las pruebas.

Recientemente se ha llegado a solicitar una pausa de seis meses a los desarrolladores de estos sistemas, hasta que se establezcan sistemas de seguridad con nuevas autoridades reguladoras, vigilancia de los sistemas de IA, técnicas que ayuden a distinguir entre lo real y lo artificial, e instituciones capaces de hacer frente a lo que algunos auguran como la "dramática perturbación económica y política (especialmente para la democracia) que causará la IA"<sup>9</sup>. Es preciso, pues, conocer los principios fundamentales del funcionamiento del sistema, ante la resistencia de las empresas de revelar detalles de los algoritmos, para ayudarnos a entender las posibles implicaciones del uso inadecuado de la IA: cámaras de eco (fenómeno que permite encontrar ideas que amplifican o refuerzan las creencias de cualquier tipo), burbujas de filtro (que supone una búsqueda personalizada, al proporcionar el algoritmo información basada en sus propios datos - localización, historial de búsquedas, páginas recientes, etc-), granjas de trols (grupo organizado de trols de internet que persiguen interferir de manera anónima en las opinión pública), noticias falsas y nuevas formas de propaganda. Venimos a identificar los riesgos del

---

<sup>6</sup> [https://www.elementsofai.com/es\\_ga=2.157383418.1937152935.1679844343-192606178.1679844343](https://www.elementsofai.com/es_ga=2.157383418.1937152935.1679844343-192606178.1679844343)

<sup>7</sup> Inteligencia Artificial y filosofía del derecho. Director Fernando H. Llano Alonso. Reflexiones sobre justicia, humanidad y digitalización. Garrido Martín Joaquín. Inteligencia (Artificial) y automatismo. Anatomía de un conflicto. Pág. 172. © Ediciones Laborum

<sup>8</sup> <https://www.dw.com/es/china-presenta-ernie-bot-el-chatbot-de-ia-que-aspira-a-ser-el-rival-más-fuerte-de-chatgpt/a-65019068>

<sup>9</sup> <https://futureoflife.org/cause-area/artificial-intelligence/>

propio funcionamiento del sistema de IA y los riesgos que entraña su utilización en Justicia.

### **II.I.- Sesgo algorítmico, discriminación.**

El sesgo cognitivo afecta a las decisiones que adoptamos consciente o inconscientemente y, sistemáticamente, provoca una distorsión de la información que se produce en el cerebro al procesarla, emitir juicios y tomar decisiones (término acuñado por los psicólogos israelíes Kahneman y Tversky). El denominado sesgo algorítmico, apunta a la incorporación al sistema de una posible discriminación por razones étnicas, de status económico, de género, de edad, demográfica, religiosas, o de otro tipo, normalmente desfavoreciendo a las minorías o grupos que no estén bien representados en los datos que se utilizarán como patrones de aprendizaje informático. Se puede producir tanto al introducir los datos o al entrenar estos dando lugar al sesgo estereotipado, introduciendo el propio prejuicio en el trabajo que realizan.

No sería más que la plasmación en el sistema del sesgo humano al recopilar datos, que o bien no son representativos de la realidad (sesgos de muestreo, cuando, por ejemplo se utilizan más rostros blancos que de otro color) o bien reflejan prejuicios ya existentes (como a título de ejemplo, el entrenamiento del algoritmo utilizando decisiones preexistentes históricas de contratación que favorecían a los hombres sobre las mujeres, manteniendo la máquina la discriminación aprendida, caso Amazon) <sup>10</sup>. Se ha constatado en los últimos años que estos sesgos tienden a repetir y perpetuar las dinámicas asentadas de dominación, privilegio y discriminación duplicando el riesgo que estas queden acentuadas y reforzadas<sup>11</sup>.

Por otro lado, al aplicar el algoritmo cálculos estadísticos sobre los datos para encontrar patrones que puedan proporcionar información, identificar tendencias y hacer predicciones, el sesgo se puede también encontrar en el aprendizaje automático, pudiendo producirse una distorsión del valor sistemático, subestimando o sobreestimando el valor de una medición<sup>12</sup> de manera que puede ofrecer resultados discriminatorios.

Si se encuentra en el momento de la preparación de los datos para seleccionar los atributos que queremos que el algoritmo tenga en cuenta, puede el algoritmo de aprendizaje automático instruirse discriminando por género, origen económico, étnico, o ideológico, pero también puede llegar a perfilar personas en función de patrones preestablecidos para seleccionar o segregar los que cumplan con aquellos. La Resolución del Parlamento Europeo sobre la inteligencia artificial en la era digital, apreció que los sesgos en los sistemas de IA, en especial en lo que respecta a los sistemas de aprendizaje

---

<sup>10</sup><https://www.technologyreview.es/s/10924/como-se-produce-el-sesgo-algoritmico-y-por-que-es-tan-dificil-detenerlo>

<sup>11</sup> Boix Palop Andrés. Transparencia en la utilización de inteligencia artificial por parte de la Administración.

<sup>12</sup> <https://revista.une.org/11/la-eliminacion-de-los-sesgos-en-los-algoritmos.html>

profundo, se producen a menudo debido a la falta de datos de entrenamiento y de prueba diversos y de alta calidad, por falta de diversidad en los equipos desarrolladores, o porque se ha comprometido el algoritmo.<sup>13</sup>

Pese a que la propuesta de regulación de la IA, prevé una serie de controles ex ante y ex post del tratamiento de los datos pretendiendo eliminar cualquier sesgo, es preciso conocer que este puede introducirse directa o sutilmente en el patrón de funcionamiento del sistema de la IA para poder valorar el resultado que ofrece.

El artículo 22 del Reglamento de Protección de Datos impide la toma de decisiones automatizadas basadas únicamente en el tratamiento automatizado que produzca un efecto jurídico o afecte significativamente de modo similar, incluida la elaboración de perfiles. Un ejemplo, podría ser la selección de candidatos a un puesto de trabajo, o el rechazo automático de una solicitud de crédito basado en el mismo. Sin embargo, la decisión adoptada en dicho perfilado solamente podría resultar ilegal, si el rechazo a que nos hemos referido se produce de modo mecánico basado en el programa sin justificación individualizada y sin intervención humana.

Con base en la normativa de protección de datos, el Tribunal Constitucional en Pleno mediante sentencia 76/2019, de 22 de mayo de 2019 en el recurso de inconstitucionalidad 1405/2019, interpuesto por el Defensor del Pueblo, declaró inconstitucional y nulo el art. 58 bis. 1 de la Ley Orgánica 5/1985, de 19 de junio, del régimen electoral general, incorporado por la Ley Orgánica 3/2018, de 5 de diciembre, de protección de datos personales y garantía de los derechos digitales, que permitía a los partidos políticos recopilar datos personales relativos a las opiniones políticas de los ciudadanos. El fallo da la razón al Defensor del Pueblo, que instó la acción para la protección de datos personales de los ciudadanos, ya que un algoritmo bien entrenado y diseñado podría, nutriéndose de los datos de una persona extraídos de los motores de búsqueda, de muestras de agrado o seguidores en redes sociales, con un grado de acierto muy elevado, conocer sus intereses, fobias y filias ideológicas, políticas o religiosas, y con ello facilitar la creación de un perfil de ideología política<sup>14</sup>.

No se ha llegado a conseguir hasta ahora probar en términos procesales la presencia del sesgo en los sistemas de IA, como en el caso Loomis, seguido ante la Corte Suprema de Wisconsin que, en el año 2016, resolvió sobre el uso del programa, COMPAS, que emplea un algoritmo predictivo de aprendizaje automático (machine learning) adoptado por la Administración de Justicia de Estados Unidos para determinar el riesgo de reincidencia del individuo en diferentes

---

<sup>13</sup> Como continúa la referida resolución, *por ejemplo, cuando se utilizan conjuntos de datos que no representan suficientemente a los grupos vulnerables, o cuando la definición de la tarea o el establecimiento de los propios requisitos están sesgados; señala que también puede producirse sesgos por la falta de diversidad en los equipos de desarrolladores de la IA, que refuerza los sesgos intrínsecos, debido al volumen limitado de datos de entrenamiento, o cuando un desarrollador de IA sesgado ha comprometido el algoritmo; señala que la diferenciación razonada también se genera de manera intencionada para mejorar el rendimiento de aprendizaje de la IA en determinadas circunstancias*

<sup>14</sup> <https://www.boe.es/buscar/doc.php?id=BOE-A-2019-9548>

etapas del proceso penal, y esa probabilidad de reincidencia afecta a la decisión sobre su puesta o no en libertad. La demanda se formuló en relación a la ausencia de respeto al "debido proceso", pero no se demostró la existencia de un sesgo por razón del origen racial de las personas en el programa, pese a que quedó acreditado que daba siempre como resultado un porcentaje de reincidencia más elevado en las personas de raza negra. Si se brindase información sobre el funcionamiento del algoritmo se podría evaluar mejor la recomendación que realiza el sistema de IA, y considerar sus limitaciones para evitar adoptar la decisión basada exclusivamente en el resultado que ofrece el sistema.

Este tipo de sistema de inteligencia artificial está categorizado en la propuesta de regulación de la Unión Europea como de alto riesgo (artículo 6.2, a), en tanto que puede tener consecuencias perjudiciales importantes para los derechos fundamentales, permitiendo su uso por parte de las autoridades encargadas de la aplicación de la ley para llevar a cabo evaluaciones de riesgos individuales de personas físicas con el objetivo de determinar el riesgo de que cometan infracciones penales o reincidan en su comisión, así como el riesgo para las potenciales víctimas de delitos, lo que supone estar sometido a controles ex ante y ex post (art.9) sin llegar a estar prohibida.

La utilización discriminatoria de los datos personales por parte del sistema fue sometida a juicio en el año 2020 en el caso SyRI, (acrónimo de sistema de indicación de riesgos), que el Gobierno holandés, utilizaba con instrumento legal para combatir el fraude de ingresos y contribuciones fiscales y leyes laborales. El sistema evaluaba el riesgo potencial que una persona pudiera hacer uso ilegal de los fondos gubernamentales, mediante el procesamiento de sus datos en dos fases: en la primera recopilaba todos los datos de la persona (becas, seguros, trabajo, sanciones, residencia, etc), tras su pseudonimización, se asignaba un número, que al cruzar con los datos de modelos y factores de riesgo, otorgaba una calificación a éste. Si el riesgo obtenido era elevado, en una segunda fase, se sometía a una unidad de análisis que califica el riesgo definitivo.

La sentencia de 5 de febrero de 2020, Rechtbank Den Haag, del Tribunal de Distrito de La Haya, entendió que el sistema algorítmico utilizado por el Gobierno de los Países Bajos para evaluar el riesgo de fraude, no cumplía las exigencias de proporcionalidad y transparencia necesarias y vulnerando las previsiones sobre respeto a la vida privada que reconoce el artículo 8 del Convenio Europeo de Derechos Humanos. Indicó el Tribunal que el procesamiento de datos afecta significativamente a una persona cuando sus efectos sean lo suficientemente importantes para afectar significativamente al comportamiento o a las decisiones de las personas involucradas; tener un efecto a largo plazo o duradero en la persona interesada; o, conducir a su exclusión o discriminación. El derecho a la privacidad de los ciudadanos fue infringido, al no hacer público el Gobierno el tipo de algoritmos utilizados en el modelo de riesgo, ni proporcionado información sobre el método de análisis de riesgos utilizado, desconociendo el afectado que dicha información era utilizada con dicha finalidad y sin poder verificar la información empleada, de modo



que al establecer perfiles de riesgo se podía producir conexiones involuntarias basadas en sesgos<sup>15</sup>.

En sentencia del 31 de diciembre de 2020, el Tribunal Ordinario de Bolonia, Italia, declaró la presencia de discriminación indirecta en las condiciones laborales de los trabajadores, en el denominado caso Deliveroo. La empresa empleaba la aplicación "Frank" que empleaba un algoritmo elaborado con las indicaciones de la empresa en cuanto a puntuación y prelación de trabajadores para que estos puedan reservar sesiones de trabajo. Dicha puntuación dependía de dos índices: fiabilidad, el número de ocasiones que el trabajador no cumple con la sesión de trabajo reservada y disponibilidad, número de ocasiones en las que el trabajador está disponible en la franja horaria de mayor demanda (20 a 22 horas de viernes a domingo). Esta aplicación discriminaba el sistema de acceso a la reserva de las franjas horarias de jornadas de los trabajadores en la plataforma, al no tener en consideración el sistema, conforme a las instrucciones de la empresa, como justificadas las ausencias de trabajo (huelga o enfermedad) incidiendo en la puntuación del rider. Ello suponía que las ausencias justificadas eran puntuadas similarmente a las no justificadas, afectando a la posibilidad de elección de horario. En la resolución no se resolvió claramente sobre el funcionamiento del algoritmo y la opacidad del mismo, pero quedó demostrado que los datos introducidos de modo sesgado por quien encomienda la elaboración del sistema, llevan a resultados discriminatorios.

En España, con la finalidad de evitar esta situación, se dictó la Ley 12/2021, de 28 de septiembre, (Ley riders) por la que se modifica el texto refundido de la Ley del Estatuto de los Trabajadores, aprobado por el Real Decreto Legislativo 2/2015, de 23 de octubre, para garantizar los derechos laborales de las personas dedicadas al reparto en el ámbito de plataformas digital como Glovo, Deliveroo o Uber Eats, que deberán adaptarse en orden a reflejar la relación laboral con sus repartidores y compartir los algoritmos, en un periodo de tres meses, obligando a la empresas a "compartir con los sindicatos las reglas en las que se basan los algoritmos que pueden afectar a las condiciones laborales", dando de dicho modo publicidad a las reglas que encierran los algoritmos y los sistemas de inteligencia artificial que pueden incidir en las condiciones laborales por las que se rigen las plataformas, incluyendo el acceso y mantenimiento del empleo y la elaboración de perfiles. Introduce una nueva letra d) en el artículo 64.4, que reconoce como derecho del trabajador a: "Ser informado por la empresa de los parámetros, reglas e instrucciones en los que se basan los algoritmos o sistemas de inteligencia artificial que afectan a la toma de decisiones que pueden incidir en las condiciones de trabajo, el acceso y mantenimiento del empleo, incluida la elaboración de perfiles".

No obstante lo expuesto, es preciso conocer que la propia tecnología ofrece sistemas para detectar y evitar los problemas del uso poco ético de la IA, aunque cumplir con la exigencia de transparencia, como veremos, mediante la explicabilidad del mismo se

---

<sup>15</sup><https://diariolaley.laleynext.es/Content/Documento.aspx?params=H4sIAAAAAAAAEAMtMSbH1czUwMDA yNDa3NDJUK0stKs7Mz7M1MjACC6rl5aekhrG425bmpaSmZealpoCUZKZVuuQnh1QWpNqmJeYUp6qJJuXnZ6OY FA8zAQCfSdKrYwAAAA==WKE>

presenta más complicado puesto que en modelos complejos de IA el efecto de cajas negras es mayor.

## **II.II.- Falta de transparencia del algoritmo, explicabilidad de la decisión.**

Arthur C. Clarke, en 1962, en su libro perfiles del futuro, formuló como una de sus tres famosas leyes la que sigue: "Cualquier tecnología suficientemente avanzada es indistinguible de la magia". Con ello podríamos describir nuestra posición ante el funcionamiento del algoritmo, a lo que ha de añadirse la presencia de lo que se denomina "caja negra"<sup>16</sup> donde opera aquel, que, en resumen, significa que ofrecen un resultado o toman una decisión sin explicar o mostrar cómo lo hicieron, al ser desconocidos los procesos internos utilizados y los múltiples factores ponderados, de modo que ni sus programadores saben cómo están tomando esas decisiones. Las entidades desarrolladoras son conscientes de ello y están creando programas que permitan observar a los usuarios del sistema de IA cómo los algoritmos toman decisiones y qué factores usan para hacer recomendaciones al final del proceso<sup>17</sup>.

Evidentemente a menor complejidad del sistema de IA, menor es la dificultad de explicabilidad del mismo. Es preciso conocer que existen también algoritmos de caja blanca en sistemas de machine learning que permiten ver el funcionamiento interno del modelo, el resultado obtenido y el árbol de decisión que detalla exactamente qué partes de la entrada se tuvieron en cuenta, cómo se ponderó cada factor, cual se ignoró, etc.

Se ha reclamado que las agencias gubernamentales dejen de usar algoritmos de caja negra<sup>18</sup>, ya que deberíamos tener mecanismos de protección equivalentes para decisiones algorítmicas como los tenemos para decisiones humanas. De modo que, para conseguir la fiabilidad del sistema y que se confíe en las decisiones autónomas, es preciso que el proceso deductivo sea inteligible y transparente, buscando necesariamente un equilibrio entre explicabilidad y complejidad, ya que como se ha indicado, a mayor complejidad menor posibilidad de explicabilidad. La transparencia de algoritmos de deep learning, por estas razones es difícil de alcanzar aunque han surgido técnicas para ello. El escollo se puede encontrar en el tipo de algoritmo al interpretar los modelos entrenados, pero se puede efectuar una clasificación digitalizada de la metodología, esto es, según se quiera realizar una interpretabilidad local - de parte del modelo- o global -

---

<sup>16</sup> La caja negra de IA hace uso de hace uso de bit data, similitud de strings (caracteres o frases), deep learning (aprendizaje profundo como parte del machine learning que permite a partir de una gran cantidad de datos y numerosas capas de procesamiento con algoritmos conseguir que el ordenador aprenda por sí mismo y realice tareas), redes neuronales (modelo computacional que emplea un conjunto de unidades -neuronas artificiales- conectadas entre sí para transmitirse señales).

<sup>17</sup> Ejemplo de ello es el programa Fairness 360 Kit. IBM que es una biblioteca de algoritmos, códigos y tutoriales innovadores que proporcionarán a los académicos, investigadores y científicos de datos herramientas y conocimientos para integrar mecanismos de detección de sesgos durante la construcción y aplicación de modelos de machine learning, identificando y reduciendo el sesgo en los modelos de IA.

<sup>18</sup> IA Now Institute de investigación que estudia las implicaciones de la concentración de poder en la industria tecnológica.

del modelo entero-; o si pretende que sirva para varios modelos o no. Es posible emplear técnicas que aproximen con un modelo explicable otro no explicable, incrementando los estudios científicos de cómo interpretar mejor los algoritmos, existiendo herramientas de código abierto que ayudan a ello<sup>19</sup>.

Este riesgo se encuentra identificado y advertido en la regulación que actualmente disponemos, ya que, el derecho a disponer de una explicación del resultado obtenido tras la aplicación de algún algoritmo de aprendizaje automático se encuentra recogido en el REGLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 27 de abril de 2016<sup>20</sup>, en tanto que el principio rector de la protección de datos consagrado en la Carta de los Derechos Fundamentales de la Unión Europea (art. 8) y que subyace en el RGPD (arts. 5.1 a) y 12), es la transparencia.

Entre las pautas éticas que han de seguirse para una IA fiable, además de atender a los principios de prevención del daño y equidad, se ha incidido en la importancia del principio de explicabilidad, como elemento crucial para construir y mantener la confianza de los usuarios en los sistemas de IA, implicando la transparencia en los procesos, capacidad y propósito de los sistemas, de manera que puedan ser explicados a aquellos quienes se van a ver afectados directa o indirectamente, por las decisiones que se adopten.

Sin embargo, como se ha apuntado, el modo en que estos datos están masivamente tratados por el sistema de IA y los factores que llevan a la toma de decisiones por los algoritmos, no son visibles, lo que conlleva la necesaria desconfianza en la calidad de la decisión. Uno de los derechos que recoge la propuesta de regulación de la IA es el de la explicabilidad en relación con los sistemas de Machine Learning. Igualmente se encuentra recogido en el artículo 15.1.h, de forma explícita en el Reglamento 2016/679 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se

---

<sup>19</sup> Mónica Vilas Olmeda. Diálogos 360, IA y Ética OdiselA [https://m.youtube.com/watch?v=x84m\\_DNfaUU](https://m.youtube.com/watch?v=x84m_DNfaUU)

<sup>20</sup> 71.-El interesado debe tener derecho a no ser objeto de una decisión, que puede incluir una medida, que evalúe aspectos personales relativos a él, y que se base únicamente en el tratamiento automatizado y produzca efectos jurídicos en él o le afecte significativamente de modo similar, como la denegación automática de una solicitud de crédito en línea o los servicios de contratación en red en los que no medie intervención humana alguna. Este tipo de tratamiento incluye la elaboración de perfiles consistente en cualquier forma de tratamiento de los datos personales que evalúe aspectos personales relativos a una persona física, en particular para analizar o predecir aspectos relacionados con el rendimiento en el trabajo, la situación económica, la salud, las preferencias o intereses personales, la fiabilidad o el comportamiento, la situación o los movimientos del interesado, en la medida en que produzca efectos jurídicos en él o le afecte significativamente de modo similar. Sin embargo, se deben permitir las decisiones basadas en tal tratamiento, incluida la elaboración de perfiles, si lo autoriza expresamente el Derecho de la Unión o de los Estados miembros aplicable al responsable del tratamiento, incluso con fines de control y prevención del fraude y la evasión fiscal, realizada de conformidad con las reglamentaciones, normas y recomendaciones de las instituciones de la Unión o de los órganos de supervisión nacionales y para garantizar la seguridad y la fiabilidad de un servicio prestado por el responsable del tratamiento, o necesario para la conclusión o ejecución de un contrato entre el interesado y un responsable del tratamiento, o en los casos en los que el interesado haya dado su consentimiento explícito. En cualquier caso, dicho tratamiento debe estar sujeto a las garantías apropiadas, entre las que se deben incluir la información específica al interesado y el derecho a obtener intervención humana, a expresar su punto de vista, a recibir una explicación de la decisión tomada después de tal evaluación y a impugnar la decisión. Tal medida no debe afectar a un menor.

deroga la Directiva 95/46/CE (Reglamento general de protección de datos) al disponer en cuanto al derecho al acceso del interesado a sus datos, debe de constar la información comprensible sobre la lógica empleada, además de la relevancia y consecuencias previstas del procesamiento automatizado de datos, siempre, y añadimos, que no afecte a otros derechos como la propiedad intelectual.

Tanto las bases de datos como los algoritmos son considerados en las empresas activos intangibles, de modo que su libre acceso puede llegar a poner en riesgo su negocio. Ello supone que, pese a la dificultad legal de su patentabilidad (RD Legislativo 1/1996, de 12 de abril), de hecho, la petición de conocimiento público del algoritmo puede ser denegada por aquellas acogiéndose a la Ley de secretos empresariales, 1/2019, 20 de febrero, permitiendo a la empresa mantener una actitud activa en el mantenimiento del secreto. Esta protección puede convertirse en un obstáculo para poder llegar a conocer la transparencia del proceso o la identificación de la parte de los elementos que, de los datos o de los algoritmos, pudieran provocar la toma de decisiones discriminatorias.

En la Carta Ética Europea sobre la IA<sup>21</sup>, se recoge el principio de transparencia, identificado como acceso al proceso de diseño completo que incluye: el diseño y la cadena operativa, como el proceso de selección, la calidad y organización de los datos que influyen directamente en la fase de aprendizaje, con código fuente abierto; la explicación del mismo mediante un lenguaje sencillo al comunicar las herramientas que se han desarrollado, el rendimiento y los riesgos de error; exigiendo que se priorice la justicia cuando la herramienta va a afectar significativamente la vida de las personas. Indica la realización de auditorías por expertos que asesoren, certifiquen y auditen los métodos de procesamiento, con una revisión periódica por las autoridades, buscando el equilibrio con la restricción que supone la protección de los secretos comerciales.

### **III.- Riesgos éticos del uso de los sistemas de IA en Justicia.**

Es endémica la lentitud atribuida a nuestro sistema judicial, debido no solo al vetusto proceso, sino también al volumen de asuntos que ingresan anualmente en nuestro país y a la ínfima ratio juez/población. Una de las posibles ventajas que, según se afirma, ofrecería el uso de sistemas de IA en justicia sería la de su agilización. También se ha apuntado como beneficio, la predictibilidad de las decisiones. Usando algoritmos se puede llegar a reducir el margen de discrecionalidad judicial, que algunos consideran excesivo, propugnando su incorporación como soporte para jueces en sus decisiones.

La Estrategia Nacional de Inteligencia Artificial (ENIA), con la intención de realizar una transición digital, impulsando la IA como elemento principal de la agenda España digital 2025, recoge como finalidad el uso de la IA para optimizar el tiempo y los recursos, y mantiene que las aportaciones de la Inteligencia Artificial al sector público que se están haciendo en numerosos ámbitos, es aplicable

---

<sup>21</sup> Carta ética europea sobre el uso de la inteligencia artificial en los sistemas judiciales y su entorno adoptado por el CEPEJ durante su 31 S t Reunión plenaria (Estrasburgo, 3-4 de diciembre de 2018)

también a la prestación de los servicios públicos en diferentes sectores de actividad entre los que incluye Justicia. Establece como línea de actuación lanzar proyectos estratégicos<sup>22</sup> para la mejora de los servicios a los ciudadanos, estableciendo un marco ético y normativo para reforzar los derechos individuales y colectivos, respetando los derechos fundamentales, la equidad en el acceso y la prevención contra la discriminación. Se prevé un "impulso normativo de revisión y reformas legislativas para la plena habilitación y operatividad del nuevo escenario tecnológico dentro de un marco democrático. Dicho impulso adquiere en la Administración de Justicia un valor estratégico añadido, si se tiene en cuenta el modelo de co-gobernanza vigente en el sector y la importancia de la interoperabilidad e interconexión de sistemas y aplicaciones".

Estos sistemas de IA están siendo utilizados en nuestro país pudiendo identificar al menos tres herramientas en las que se emplean algoritmos predictivos: la valoración policial del riesgo de reincidencia en violencia de género, sistema Viogen; la gestión penitenciaria en las prisiones de Cataluña, sistema Rise Candy y la valoración del riesgo de radicalismo violento en prisión. Aunque en puridad no puedan ser calificados como IA, por cuanto que solo utilizan grandes cantidades de datos para obtener patrones sin aprendizaje profundo, sugiriendo la doctrina más autorizada<sup>23</sup> la utilización de estos instrumentos como herramienta auxiliar o complementaria en la toma de decisiones penales, a la hora de imponer una prisión provisional, sin que la calificación del riesgo sea vinculante. Y como herramientas auxiliares presentan unas ventajas incontestables, como apunta Antonio del Moral, al referirse a un instrumento puesto a disposición de los magistrados, la calculadora 988, que, en segundos, realiza las tareas de acumulación de penas según los criterios del art. 76 del Código Penal, abogando por la utilización de sistemas de IA en decisiones protocolizadas en la investigación y en la tramitación, aunque recuerda que en la tarea estricta de enjuiciar, precisar qué es lo justo, lo de cada uno -y eso es el núcleo duro de la función judicial- el factor humano es irrenunciable e imprescindible.

En otros países, como en el Reino Unido<sup>24</sup>, la Corte inglesa respaldó expresamente el uso de software de codificación predictiva, e-discovery, que identifica, recolecta y examina documentos almacenados en un sistema para su utilización en la fase procesal en que se ha de proceder a la divulgación de los documentos que dispone (disclosure) al exigirse a cada parte que reúna e identifique los documentos potencialmente importantes para el caso y, que señale, posteriormente, el resto de documentos que existen o hayan existido con anterioridad, conociéndose de ese modo desde el primer momento la

---

<sup>22</sup> LÍNEA DE ACTUACIÓN 5.2. LANZAR PROYECTOS ESTRATÉGICOS RELACIONADAS CON LA IA MEDIDA 25. Promover misiones estratégicas nacionales en el ámbito de la administración pública donde la IA puede tener impacto para la mejora de los servicios a los ciudadanos (áreas tales como salud, justicia, empleo, etc.

<sup>23</sup> Martínez Garay Lucía. Paradojas de los algoritmos, predictivos utilizados en el sistema de justicia penal. Revista El cronista del Estado social y Democrático de derecho. número 100.

<sup>24</sup> Caso Pyrrho Investments Ltd vs. MWB Property Ltd (2016)

documentación existente y que conlleva que la otra parte tenga derecho a una copia. La decisión del Tribunal se basó en varios argumentos, a saber: que varias jurisdicciones probaron que dicho software fue útil en los casos apropiados, que las partes acordaron utilizar el software, que no hubo evidencia que sugiera un menor grado de precisión, o una alternativa más consistente y barata; y finalmente, no hay una legislación que prohíba su uso.

Ejemplos de uso de la IA en el Poder Judicial de otros países los encontramos en India, donde se está utilizando tribunales en línea para resolución de procesos relativos a indemnizaciones derivadas de siniestros de circulación y otras reclamaciones menores; emplean también un software que se basa en el procesamiento del idioma natural para traducir resoluciones y sentencias a idiomas regionales y otros sistemas de IA para datos de investigación jurídica y predecir el progreso de un caso. En Brasil, la herramienta llamada VICTOR, introducida a iniciativa del Tribunal Supremo del país, lee los recursos de apelación que llegan hasta el Supremo Tribunal en su fase de admisión e identifica cuáles están enlazadas a ciertos temas de repercusión general; en Pernambuco, un sistema de IA examina las nuevas acciones de ejecución fiscal y decide cuáles están de acuerdo con los reglamentos de los procedimientos, y cuáles serán desestimadas debido al régimen de prescripción; y los Tribunales Federales de Apelación para la tercera Región comenzaron la implementación del programa de IA llamado SIGMA, para ayudar en la preparación de informes, resoluciones y sentencias en el sistema de Proceso Judicial Electrónico. En Noruega, la Comisión de la Administración Pública emitió un informe en 2019 recomendando la toma de decisión automatizada en los procedimientos administrativos judiciales para mejorar la eficiencia, en particular en donde el volumen de casos es grande. Por último, en Estonia existe un programa piloto de juez de IA "juez robot" que pueda decidir litigios de menor cuantía (inferior a 7.000 euros), funcionando de modo que ambas partes suben documentos y otra información relevante a la plataforma del tribunal y en función de dichos datos la IA emite una resolución que pueda ser apelada ante un juez humano.<sup>25</sup>

Sin poner en duda la efectividad de dichos sistemas y la gran labor de auxilio en el desempeño de la función judicial que pudieran ofrecer, debiéramos cuestionarnos si, aun cuando la última decisión se residencie en el humano, esta no estaría condicionada por la adoptada previamente por la máquina, provocando una suerte de autosabotaje en el proceso deductivo, dejando de atender a la intuición, la experiencia o la apreciación personal de las características del caso, que llevasen a una conclusión distinta de la máquina. Partiendo de la utilización presente (en la fase policial, de investigación, para la predicción de resultados, etc) y futura de los sistemas de IA,

---

<sup>25</sup> La IA y el Estado de derecho. Curso Unesco y The National Judicial College. Defender el Estado de derecho en la era de la IA. *Estonia está entre los países mejor clasificados en el Índice de Desarrollo de Gobierno Electrónico en 2018 y 2020, ya que sus ciudadanos y servidores públicos pueden acceder a una amplia variedad de servicios en línea usando ID digitales y seguras, incluyendo hacer pagos, acceder a registros completos de salud y votar por Internet. Hoy, 99% de los servicios públicos están disponibles en línea 24/7, 30% de los estonios usan el voto por Internet (i-Voting), y los estimados de la reducción de la burocracia ha ahorrado 800 años de tiempo de trabajo. Hasta ahora, las demandas de menor cuantía han sido delegadas a un tribunal en línea que funciona exclusivamente con IA*

debiéramos diferenciar conceptos Justicia y Administración de Justicia para identificar los peligros del uso de aquella en ambos.

Si nos referimos a la Administración de Justicia, entendida como gestión de esta a través de un sistema, atendiendo a la natural tensión certidumbre-artificial/incertidumbre-humana, se considera por la mayoría de la doctrina, que puede ser una oportunidad dentro del ámbito de la Administración de justicia pero manteniendo siempre, la dimensión instrumental de la tecnología<sup>26</sup>. Pero paralelamente también puede presentar riesgos que pasamos a identificar.

### **III.I.- Uso de IA en el proceso de toma de decisiones**

El proyecto de Reglamento Europeo considera como de alto riesgo ciertos sistemas de IA destinados a la administración de justicia, dado que pueden tener efectos potencialmente importantes para la democracia, el Estado de Derecho, las libertades individuales y el derecho a la tutela judicial efectiva y a un juez imparcial. A fin de evitar el riesgo de posibles sesgos, errores y opacidades, categoriza como tales aquellos sistemas de IA cuyo objetivo es ayudar a las autoridades judiciales a investigar e interpretar los hechos y el Derecho y a aplicar la ley a unos hechos concretos, sin embargo, no extiende dicha clasificación a los sistemas de IA destinados a actividades administrativas meramente accesorias que no afectan a la administración de justicia en casos concretos, como las tareas administrativas, o la asignación de recursos<sup>27</sup>.

De modo que distingue, por la potencial afectación a las libertades individuales y derechos fundamentales, en cuanto a la asignación del riesgo, entre lo que define como tareas simplemente administrativas y las que conlleven aplicación o interpretación del derecho, lo que a nuestros efectos, lleva a plantearnos qué podemos considerar como tareas en el seno de la tramitación de un asunto que puedan integrar el primer concepto. ¿Las que se encuentran exentas de ámbito de decisión jurisdiccional, el mero impulso del proceso, la decisión tomada sin margen discrecional, o cualquier otra que esté excluida de la intervención judicial aunque esté sometida a su revisión?.

El Proyecto de Ley de Medidas de Eficiencia Digital del Servicio Público de Justicia de 22 de septiembre de 2022, apuesta por la utilización de lo que denomina el dato abierto, para la toma de decisiones estratégicas en procedimientos concursales, los juicios de desahucio o los procedimientos en materia de violencia sobre la mujer. Prevé la implantación de lo que denomina actuaciones automatizadas en el seno del proceso, que define como aquellas de carácter procesal que no requieren intervención humana para cada caso singular <sup>28</sup> ,

---

<sup>26</sup> Inteligencia Artificial y filosofía del derecho. Director Fernando H. Llano Alonso. Reflexiones sobre justicia, humanidad y digitalización. Stefano Bini. Pág. 79. © Ediciones Laborum

<sup>27</sup> Considerando 40 del Proyecto de Reglamento de IA

<sup>28</sup> Art. 56. Y define las actuaciones productivas automatizadas aquellas que sean auto iniciadas en relación con los sistemas de información sin intervención humana y que aproveche la información incorporada en un expediente o procedimiento de una administración pública para un fin determinado, generando avisos o

autorizando la automatización de las actuaciones de trámite o resolutoria simple, definiendo estas con las que no requieren interpretación jurídica. Incluye como posible, la utilización de sistema de IA para una actuación asistida, que define como el sistema que genera un borrador total o parcial de un documento complejo basado en datos, producido por algoritmos y que puede constituir fundamento o apoyo de la resolución judicial o procesal<sup>29</sup>, sin que constituya una resolución per se, sin validación.

Nuestro compañero Javier Ercilla recientemente ha publicado un artículo, donde plantea el uso del programa desarrollado utilizando el motor de inteligencia artificial GPT-4 de OpenIA (imagínate es un sistema de inteligencia artificial de la categoría de los procesamientos de lenguaje natural lanzado el 15 de marzo de 2023 que permite generar respuestas fundamentadas y coherentes a las cuestiones que se les somete) como herramienta útil para resolver revisiones fácticas en recurso de suplicación en el ámbito social. Según el compañero las pruebas realizadas con el sistema de IA acreditan que es capaz de generar respuesta adecuada, coherente con el contexto proporcionado, resolviendo correctamente la mayoría de los recursos de suplicación presentados. Ahora bien también identificó dificultades para comprender completamente el contexto o aplicar correctamente las directrices legales que le eran dadas, sin que pueda realizar tareas de cognición profunda en el sentido de procesos cognitivos más complejos, ya que lo único que pueden es imitar ciertos aspectos de estos procesos, porque su comprensión se basa en patrones y

---

efectos directos a otros fines distintos en el mismo o en otros expedientes, de la misma o de otra administración pública, en todo caso conformes con la Ley (art. 56.2).

<sup>29</sup> Artículo 57. Actuaciones asistidas.

1. Se considera actuación asistida aquella para la que el sistema de información de la Administración de Justicia genera un borrador total o parcial de documento complejo basado en datos, que puede ser producido por algoritmos, y puede constituir fundamento o apoyo de una resolución judicial o procesal.

2. En ningún caso el borrador documental así generado constituirá por sí una resolución judicial o procesal, sin validación de la autoridad competente. Los sistemas de la Administración de Justicia asegurarán que el borrador documental sólo se genere a voluntad del usuario y pueda ser libre y enteramente modificado por éste.

3. La constitución de resolución judicial o procesal requerirá siempre la validación del texto definitivo, por el juez o jueza, magistrado o magistrada, fiscal o letrado o letrada de la Administración de Justicia, en el ámbito de sus respectivas competencias y bajo su responsabilidad, así como la identificación, autenticación o firma electrónica que en cada caso prevea la ley, además de los requisitos que las leyes procesales establezcan.

Artículo 58. Requisitos comunes de las actuaciones automatizadas, proactivas y asistidas.

1. En caso de actuación automatizada, asistida o proactiva podrá realizarse por el Comité técnico estatal de la Administración judicial electrónica la definición de las especificaciones, programación, mantenimiento, supervisión y control de calidad y, en su caso, la auditoría del sistema de información y de su código fuente.

2. Los criterios de decisión serán públicos y objetivos, dejando constancia de las decisiones tomadas en cada momento.

3. Los sistemas incluirán los indicadores de gestión que se establezcan por la Comisión Nacional de Estadística Judicial y el Comité técnico estatal de la Administración judicial electrónica, cada uno en el ámbito de sus competencias.



correlaciones aprendidos a partir de los datos con los que han sido entrenados. Lo considera válido como una herramienta de apoyo, pero no como una solución definitiva, siempre bajo el control del profesional y la asunción de la responsabilidad en la toma de decisiones, en palabras del propio Ercilla, "como un dictaminador de posibilidades verificables por el usuario, más que como sistema de decisión final". Finalmente plantea que su implementación provocaría cambios en el rol de los profesionales del derecho ya que requeriría que estos adquirieran habilidades en el uso de estas tecnologías y pudiendo provocar un cambio en la toma de decisiones puesto que los jueces pudieran verse influenciados por las soluciones que propone GPT-4, ello sin contar con los problemas que pudieran surgir de la presencia de sesgos en el entrenamiento del programa<sup>30</sup>.

Llegados a este punto, sabiendo que ya se están empleando y que se prevé en un futuro cercano la utilización sistemas de IA durante determinadas fases del proceso que concluye con la toma de decisión por el juez, es el momento de valorar su utilización durante el iter procedimental y cómo auxiliar en la propia toma de decisión, en tanto que puede afectar a derechos reconocidos o al propio orden constitucional. Su uso potencialmente puede afectar, entre otros, a derechos fundamentales como el derecho a la intimidad -protección de datos- o a la tutela judicial efectiva, pero también puede ofrecer indudables ventajas.

### **III.I.I.- Seguridad jurídica, imparcialidad, derecho al juez natural.**

La seguridad jurídica, entendida como predictibilidad de las consecuencias jurídicas de los actos o conductas y como Principio General del Derecho, en cuanto que condiciona la actuación de los aplicadores del derecho en relación con un conjunto indeterminado de supuestos semejantes, ya venía recogida en la Declaración de Derechos del Hombre y del Ciudadano de 26 de agosto de 1789 (sûreté) y en nuestra Constitución Española, Título Preliminar.<sup>31</sup> Indudablemente el uso de sistemas de IA garantizaría el cumplimiento de este principio, y apuntamos, que ello siempre que fuese totalmente posible explicar la decisión tomada por el sistema, ya que en caso contrario, cederíamos la obligada argumentación jurídica ciegamente por la resolución adoptada por la máquina, que necesariamente exigiría de una explicación pero a posteriori.

Se deduce de lo anterior, que el principal riesgo en el uso de esta tecnología en el ámbito de la justicia, lo encontraríamos en el hecho que estos sistemas, que, como hemos expuesto, ya de por sí pueden presentar disfunciones en cuanto a los resultados obtenidos derivados de la falta de transparencia de la decisión o bien de la

---

<sup>30</sup> Ercilla García Javier. Inteligencia artificial y revisión fáctica en la jurisdicción social: integrando GPT-4 en la resolución de los recursos de suplicación. Revista APM 3.9

<sup>31</sup> Artículo 9.3, según el cual "la Constitución garantiza el principio de legalidad, la jerarquía normativa, la publicidad de las normas, la irretroactividad de las disposiciones sancionadoras no favorables o restrictivas de derechos individuales, la seguridad jurídica, la responsabilidad y la interdicción de la arbitrariedad de los poderes públicos"

propia presencia de sesgo en el tratamiento de los datos, estos se trasladen a la decisión última que se adopte con base a estos sistemas, lo que afectaría necesariamente al derecho que todos los ciudadanos tienen obtener una resolución dictada por un juez imparcial, al venir afectada dicha imparcialidad de origen<sup>32</sup>. Sin embargo, como medio auxiliar del juez, poniendo a su disposición los precedentes, criterios o posibles soluciones al supuesto sometido a su decisión, permitiría agilizar y mejorar la decisión al reducir el tiempo de búsqueda, permitir el contraste con su propio conocimiento y perspectiva en la toma de decisión con la que le ofrece la máquina.

Existiendo experiencia en otros países de nuestro entorno<sup>33</sup>, comprobándose un aumento de las medidas alternativas de solución de disputas en línea (ODR)<sup>34</sup> y su posible extensión a los denominados tribunales en línea para disputas de baja intensidad, se está planteando admitir la sustitución de la decisión del juez por el sistema de IA en los casos sencillos<sup>35</sup>, ahora bien, sería precisa una definición clara, taxativa y consensuada de lo que ha de entenderse por tal. Conociendo, como sabemos, que no hay dos casos idénticos, presentando matices que la máquina no podrá advertir, dichos matices podrían ser introducidos en la resolución con la revisión posterior. En todo caso, aunque, consideráramos como tales, aquellos casos que cumplan con unos parámetros preestablecidos sin margen de discrecionalidad derivado bien de la preexistencia de precedentes (línea jurisprudencial firme), bien por la aplicación automática de disposiciones legales (normas imperativas), se estaría cediendo la función jurisdiccional de juzgar constitucionalmente atribuida al poder Judicial al poder gubernativo que encargara o desarrollara el sistema de IA decisor. Resultaría por tanto, para ello preciso en nuestro país una modificación constitucional que avalase tal decisión al afectar a la previsión del art. 117 de la CE, y por ende al principio de separación de poderes. Si bien, se podría salvar tal menoscabo de la función constitucional mediante la asunción judicial del control en la selección de datos, desarrollo y en la implementación de los sistemas de IA en justicia.

Por otro lado, hay que prestar especial atención a que la utilización del sistema de IA pudiera llegar a afectar al derecho al juez natural, al juez ordinario predeterminado por la ley. Podría pensarse que la implementación de este tipo de programa para el reparto de asuntos judiciales es una tarea administrativa, simplemente mecánica, que se encuentra fuera de los peligros que se atisban por el uso de IA. Sin embargo, ya ha habido suspicacias que han desembocado en una resolución, al emplear, la reciente reforma judicial en Polonia (2018), el denominado Sistema de Asignación Aleatoria de Casos ("SLPS"), que decide cómo asignar casos judiciales a los jueces en los

---

<sup>32</sup> Inteligencia Artificial y Jurisdicción Penal. Discurso de ingreso como Académico de Número, en la Real Academia de Doctores. Marchena Gómez Manuel. pág. 47

<sup>33</sup> Gran Bretaña, los Países Bajos y Letonia

<sup>34</sup> Servicios de conciliación, mediación y arbitraje fuera de la sala del tribunal

<sup>35</sup> Inteligencia Artificial y filosofía del derecho. Director Fernando H. Llano Alonso. Reflexiones sobre justicia, humanidad y digitalización. La Justicia Predictiva: tres posibles usos de práctica jurídica. De Asís Pulido Miguel. pág. 298.

tribunales, generando dudas sobre su carácter azaroso, y que dio lugar a la interposición de una demanda reclamando de la Administración revelar el algoritmo al ostentar este carácter. No puede considerarse como información técnica en exclusiva ya que es una expresión de un procedimiento estrechamente relacionado con la situación directa de los ciudadanos cuyos casos se examinan, debiendo describirse paso a paso la operación del SLPS, que se ofreció. Como consecuencia de dicha resolución, ha descrito el Ministerio verbal y gráficamente cómo funciona el algoritmo en varias etapas del del sistema SLPS, indicando que debe de estar abierto y disponible públicamente tanto el algoritmo como los parámetros de entrada, por ejemplo, una lista de jueces, sus coeficientes de asignación, etc.

Sin embargo, pese a lo anterior, la cognoscibilidad total no se ha logrado, porque si sólo se publica el código fuente, también podrían generarse dudas sobre la aleatoriedad del generador de números que asignen casos a determinados jueces, o sobre la posible modificación sutil del sistema con exclusiones del sorteo, de modo que solo la total transparencia del funcionamiento generaría la confianza en el mismo y el control social de las instituciones públicas<sup>36</sup>.

### **III.I.I.- Opacidad de las propuestas.**

Resultaría para la función de juzgar extremadamente peligroso desconocer que se están empleando sistemas de IA en la propuesta de resolución que se ofrece, como la falta de posibilidad de explicar la lógica que subyace en la decisión adoptada con base a un proyecto propuesto por un sistema de IA, sustrayendo al afectado del conocimiento que sus derechos han sido respetados durante todo el proceso de desarrollo del algoritmo<sup>37</sup>. Ahí reside la necesidad de transparencia total del funcionamiento en el caso concreto del algoritmo que, en ningún caso, en Justicia, pueden ser empleados los de caja negra, al no poder explicar la decisión. La adoptada debe estar debidamente justificada en relación a los datos ofrecidos para el caso concreto y las posibles soluciones que pudiera haber adoptado, sin que pueda asumirse de modo acrítico la propuesta por el sistema, aun cuando exista acreditación de haber respetado los derechos fundamentales y los principios éticos durante su desarrollo.

Otro de los problemas que se podrían atisbar del uso de los sistemas de IA al tiempo de resolver las cuestiones que se le sometan a debate, es la dificultad que se presenta a la parte que quisiera formular recurso discrepando de la calificación jurídica ofrecida por el sistema, ya que, si el algoritmo ha sido programado de manera que haya tenido en consideración la jurisprudencia existente sobre el supuesto, y de los preceptos aplicados en anteriores precedentes, resultaría difícil encontrar el argumento jurídico en que podría fundarse la pretensión revocatoria, lo que provocaría que habría de enfocar los esfuerzos demostrativos, más en el hecho de la ausencia de datos que hubo tenido en consideración al algoritmo y en la

---

<sup>36</sup> SLPS-access-to-algorithm-ruling.docx 28.77 KB

<sup>37</sup> Boix Palop Andrés. Transparencia en la utilización de inteligencia artificial por parte de la Administración. El cronista del Estado Social y democrático de derecho. Núm. 100. Pág. 92-96

modificación del resultado que se produciría de haberlos incluido, que en un argumento jurídico, dando paso un razonamiento más técnico-informático que técnico-jurídico<sup>38</sup>.

Por otro lado, el razonamiento judicial que contiene la resolución dictada, se funda - y así se explica en la misma - en la identificación y valoración de los hechos relevantes, a los que posteriormente sigue una declaración de los que han sido o no probados, aplicación de normas legales y jurisprudencia que la interpreta, para llegar a la solución que se ha considerado adecuada a ese caso concreto, coherente con el razonamiento que expone, partiendo de la premisa que, aun conociendo que no es la única solución correcta, sí permite justificar, argumentando jurídicamente, el motivo de la opción escogida.

Este trabajo de interpretación no lo hacen las técnicas de aprendizaje automático. No pueden elaborar un razonamiento legal ya que su opción se basa en parámetros de uniformidad al procesar grandes cantidades de información, obteniendo una solución que según sus rígidos parámetros podría ofrecer un juez en esta situación. No existe ningún tipo de razonamiento jurídico aplicado al caso concreto, ni tampoco comprende, y por tanto no puede justificar, la motivación de la adopción de esa decisión y no otra.

Refiere la citada Carta Ética Europea sobre IA, una experiencia llevada a cabo en Francia a iniciativa del Ministerio de Justicia. Los dos tribunales de apelación de Rennes y Douai acordaron probar el software de justicia predictiva en varios recursos judiciales realizando un análisis de las decisiones civiles, sociales y comerciales de todos los tribunales franceses de apelación, un análisis cuantificado de las sumas otorgadas por estos dos tribunales, y una clasificación geográfica de las discrepancias observadas para juicios similares. La finalidad del software era crear una herramienta de toma de decisiones para reducir, si es necesario, la variabilidad excesiva en las decisiones judiciales. El resultado obtenido, debatido por el Ministerio y la empresa que diseñó el producto, fue la ausencia de valor agregado del sistema para el trabajo de toma de decisión y reflexión de los magistrados, al revelarse sesgos del razonamiento del software que condujeron a resultados aberrantes o inapropiados debido a la confusión en algún caso meramente léxico del razonamiento judicial y las causas que habían sido decisivas en el razonamiento de los jueces.

Añadir, por último, el riesgo advertido que las decisiones que se adopten por la IA no sean aceptadas y que exijan un exceso de argumentación para basar una decisión que un principio habría de ser adoptada de modo normal<sup>39</sup>

---

<sup>38</sup> Marchena Gómez Manuel. Opus. Cio. Pág. 41

<sup>39</sup> Inteligencia artificial y Filosofía de derecho. La problemática de los sesgos algorítmicos (con especial referencia a los de género). ¿Hacia un derecho a la protección contra los sesgos? Beloso Martín Nuria. Pág 55. @edicioneslaborum

### **III.I.III.- Eficientismo. Perfilado de jueces.**

La búsqueda permanente de un proceso legal más eficiente del que actualmente disponemos ha generado la expectativa de conseguir alcanzar dicha meta mediante el uso de herramientas de IA y a dicha aspiración quiere dar respuesta el proyecto de ley de medidas de eficiencia digital del servicio público de Justicia que la considera una demanda real y urgente de la ciudadanía<sup>40</sup>.

Pese a la bondad de la aspiración, se atisba el riesgo que, bajo la justificación de la eficiencia, se pretenda extender dicha utilización a aspectos que deben de estar controlados por el ser humano, en concreto por el juez. Puede cederse a la tentación de dirigir la labor judicial, limitarla en alguna medida, mediante la utilización de programas informáticos que, al introducir datos o al entrenar al sistema, comprometan la neutralidad del mismo y que, por resultar su empleo imperativo, no permitan modificar el resultado que ofrecen so pretexto de inexistencia de margen alguno de interpretación jurídica. Ello afectaría gravemente la independencia del poder judicial ya que, podrían adoptarse decisiones que condicionaran la respuesta judicial bajo el pretexto de una ficticia objetivación en la toma de aquellas, mejora en la eficiencia de la administración de justicia y agilización de la tramitación del procedimiento.

Se puede conseguir emplear una IA responsable en justicia, basándose en una potenciación de los aspectos positivos y una correlativa mitigación de los negativos, ajustándola a unos principios éticos que se están universalizando mediante un consenso de mínimos, empleando sistemas que tengan asegurada la ética por diseño, procurando que el propio sistema ofrezca los medios para detectar y eliminar utilizando sistemas para ello los riesgos de disfunciones de la IA<sup>41</sup>.

No pueden emplearse los sistemas de IA como instrumento cercenador de, lo que algunos consideran, el excesivo margen de discrecionalidad judicial propugnando, incluso, el uso de herramientas de perfilado de jueces y magistrados para crear en el imaginario colectivo la idea que el empleo del sistema de IA eliminaría, o al menos controlaría, los sesgos cognitivos de las decisiones. Y no puede ser ese su objetivo, ya que se obvia la posible presencia de aquellos en el propio sistema de IA, provocando la sustitución de la función de juzgar constitucionalmente atribuida a los miembros del poder judicial, por la decisión administrativa basada en los patrones que decida introducir la Administración. Se ha indicado que la ética de la IA comienza con la ética de los propios programadores, y añadiría de los que encomiendan su desarrollo. Debiendo de advertirse, como hace la Comisión Europea para la Eficiencia de la Justicia (CEPEJ), que los riesgos de interpretaciones distorsionadas del significado de las decisiones judiciales son extremadamente altos cuando se basan

---

<sup>40</sup> [https://www.congreso.es/public\\_oficiales/L14/CONG/BOCG/A/BOCG-14-A-116-1.PDF](https://www.congreso.es/public_oficiales/L14/CONG/BOCG/A/BOCG-14-A-116-1.PDF)

<sup>41</sup> Encriptación homomórfica - técnica que permite realizar operaciones sobre los datos cifrados y obtener resultados cifrados equivalentes a las operaciones realizadas sobre la información original-, auto-machine Learning, disposición pública del sistema de gobernanza y escalado, etc.

únicamente en modelos estadísticos<sup>42</sup>, de modo que, en ausencia de una representación estadística de la realidad o de poder predecir algo, los resultados del software de justicia predictiva se establecerán como estándares sin ninguna validación por parte del sistema legal y en conflicto con él<sup>43</sup>.

A lo anterior se añade la deriva totalitaria que supondría permitir al sistema elaborar un perfilado de los jueces que emplee indagaciones, mediante el cruzado de datos, con las propias redes sociales, sitios comerciales, búsquedas etc, bien para indagar sobre supuestos sesgos ideológicos que pudieran apreciarse en la respuesta a la cuestión que se le somete a debate - cuando se ha advertido por la experiencia que ha sido el uso de algoritmos fallidos lo que ha provocado situaciones discriminatorias<sup>44</sup>, o incluso para evaluar el desempeño de los mismos.

Supondría permitir tal uso una vulneración de los datos personales del juez, ya que la exigencia de la mención del nombre y apellidos de quien dicta la resolución conforme al principio de juicio público establecido en el artículo 6 del Convenio Europeo de Derechos Humanos, que garantiza la imparcialidad y el cumplimiento de las normas procesales, no puede amparar la utilización de estos para realizar cálculos estadísticos sobre la ideología, religión, creencias de un juez en particular, atentando<sup>45</sup> contra su libertad ideológica, religiosa o de culto<sup>46</sup>. El ejercicio de la función judicial debe ser juzgado por el contenido de las resoluciones, no someter a un juicio al juzgador quien, a su vez, no debe juzgar personas sino hechos<sup>47</sup>

Por otro lado, es fácilmente apreciable que verse sometido a dicha evaluación podría afectar a la libertad de criterio de éste, por la presión a la que pudiera verse sometido al tiempo de valorar la decisión a adoptar ya que vendría condicionado a resolver procurando evitar la lectura de su resolución bajo la perspectiva sesgada que le atribuye el sistema. No es aventurado afirmar que puede incluso ceder a la presión de actuar en función de la predicción o mimetizar la resolución ajustándola a línea fijada por la estandarización de las decisiones. A ello se añade el riesgo cierto del empleo de sistemas de IA que efectúen un perfilado de los jueces ante los que se va a actuar, permitiendo a las empresas privadas y sus abogados participar de prácticas de compra en el foro, otorgando una ventaja tan

---

<sup>42</sup> Carta ética de la IA en los sistemas de justicia. Apart. 104

<sup>43</sup> Carta ética de la IA en los sistemas de justicia. Apart. 149

<sup>44</sup> Inteligencia artificial y Administración de Justicia: ¿Quo vadis, justitia?. Simón Castellano Pere. Universidad Internacional de La Rioja, <https://dialnet.unirioja.es/download/articulo/7962069.pdf>

<sup>45</sup> STC, 120/1990 de 27 junio.

<sup>46</sup> Si bien se han indicado limitaciones a la libertad ideológica del juez - JORGE F. MALEM SEÑA, Los jueces: ideología, política y vida privada.- sin embargo los propuestos no suponen la anulación del derecho, sino que se encuentran íntimamente imbricados con la función judicial.

<sup>47</sup> Diez reglas de deontología judicial, consecuentes a la naturaleza cognoscitiva de la jurisdicción. Luigi FERRAJOLI. Séptima regla. Nolite iudicare. Juicios sobre hechos no sobre personas.

competitiva, que se vea socavado el principio de un juicio justo establecido por ley<sup>48</sup>.

En Francia, imponiendo penas de cárcel de hasta cinco años, se ha prohibido la publicación de información estadística sobre decisiones y el patrón de conducta de los jueces en cuanto a las sentencias que dictan, limitando la funcionalidad de herramientas de Legal tech o como la francesa Predictice o la española Jurimetría. El propósito de impedir el empleo de los datos de identidad de los magistrados, es evitar presiones y mercadeo de estrategias procesales; sin embargo esta decisión no está exenta de polémica ya que se ha manifestado que dichos datos pueden extraerse fácilmente de las resoluciones dictadas en cuanto que sean públicas, y podría ser considerada como una contribución a la satisfacción de los principios propios de una organización política democrática, especialmente el principio de igualdad vinculado al de seguridad jurídica (artículo 14 de la CE en relación con el 9.3)<sup>49</sup>.

Resulta aterrador imaginar que pudiera implantarse en nuestros sistemas judiciales sistemas de IA para llevar a cabo una auditoría de las decisiones judiciales, con independencia de la bondad que se atribuya a la finalidad de su adopción, ya que, dejar en manos de otro poder la valoración de la decisión que habría de adoptarse, imponiendo al juez adaptarse a las decisiones que considere que hayan de ser estandarizadas, rompe completamente con el principio de separación de poderes y la independencia del poder Judicial colocando a este en una posición subordinada a disposición del gobierno de turno. No se plantea como un supuesto de laboratorio, es una realidad, siguiendo el modelo chino donde existe un aviso de sentencias "fuera de lo normal" alertando a las autoridades cuando se ha excedido de los límites de discrecionalidad establecidos en precedentes, se propugna su utilización para detectar patrones discriminatorios del sistema judicial, e incluso "influencias emocionales excesivas, influjos ideológicos inapropiados o el uso acrítico de heurísticos del pensamiento"<sup>50</sup>

### **III.II.- Ética de los sistemas IA en Justicia**

Si la ética pretende alcanzar la adecuación el comportamiento humano a los principios básicos de beneficencia, no maleficencia, respeto por las personas y justicia, cuando se reclama que la IA ha de ser ética, se refiere a la necesidad que su diseño y despliegue no encubra comportamientos que humanamente consideramos reprochables. La Justicia entendida como concepto filosófico, como virtud, definida por Aristóteles, en el modo de ser por el cual se está dispuesto a practicar lo que es justo, obrar justamente y a querer lo justo (Ética

---

<sup>48</sup> Como ha puesto de relieve la CEPEJ en la referida Carta ética europea sobre el uso de la IA en los sistemas judiciales y su entorno: Esta táctica ya se ha observado durante mucho tiempo en los Estados Unidos y en Francia por delitos de prensa y violaciones de la privacidad en la prensa, donde ya se sabe que los demandantes eligen el tribunal que parece otorgar la mayor cantidad de daños e intereses.

<sup>49</sup> Justicia Digital. Guía para el diálogo sobre el diseño y uso eficiente, de calidad y ético de herramientas tecnológicas en la justicia civil. Cotec, Universidad de Granada, Sociedad Digital.

<sup>50</sup> De Asís Pulido Miguel. Opus cit. 303 – 305

Nicómaquea) o con Platón que la entiende como poseer y hacer lo que es propio, obrar justo, no hacer daño a nadie (libro segundo de La República), no puede actualmente ser impartida por una máquina, ya que se sabe que estamos muy lejos de crear la que se denomina IA general, fuerte, en el sentido que sea capaz de generar sus propios objetivos. Se ha intentado antropomofizar la IA, afirmando que la general puede llegar a tener consciencia, sin embargo los expertos aseguran que no estamos ni siquiera cerca de hablar de conciencia en una IA<sup>51</sup>, ya que las tecnologías actuales, como DALL-E2 , GPT-4 y demás, siguen estando muy lejos de alimentar lo que imaginamos como una máquina consciente, por espectaculares que sean sus resultados actuales.

Como el sistema carece de consciencia, inteligencia o contexto, evidentemente no puede ser capaz de aprehender el significado de lo que es justo ni la distinción de lo que no lo es. Markus Gabriel, considera que los sistemas descritos como inteligencia artificial "no presentan, en realidad, una forma de inteligencia ni la capacidad de pensar, ya que carecen de contexto y por lo tanto de sentido. Son más bien modelos de pensamiento, imágenes de la manera en que pensamos. No existe una inteligencia artificial real, es pura especulación, pero su imposición discursiva es importante: permite la dominación con la tecnología y cumple un rol deshumanizador". En el mismo sentido se expresa Byung- Chul Han, ya que al actuar exclusivamente mediante adición y acumulación de información digital, fragmenta la vida porque carece de memoria e historia, desbancando la narración, que es lo que crea significado y contexto, por tanto al faltarle el concepto, no comprende los resultados de sus cálculos<sup>52</sup>.

En definitiva, siguiendo a estos filósofos actuales, con la utilización de patrones con los que dar respuesta uniforme a problemas similares (algoritmo), todo se vuelve predecible y controlable porque se cuenta con datos preexistentes, pero no supone ningún avance. No permite un razonamiento ante imprevistos o con circunstancias alteradas, de modo que, como indica Byung-Chul Han, es una forma de saber primitiva ya que maneja datos preexistentes pero no razona, únicamente establece una correlación<sup>53</sup>.

El nihilismo - reinventado en el siglo XIX - que impera en nuestra sociedad, sustentado en esta tecnología, puede provocar el riesgo de permear en concepciones jurídicas reconocidas como esenciales, eliminando la importancia del logro que supuso la universalización del reconocimiento de los derechos humanos, cediendo ámbitos de consenso mundial en pos del uso de una IA, carente de humanismo, olvidando que la historia no nos lleva automáticamente a un progreso jurídico-moral <sup>54</sup> . Si se consiguiese alcanzar lo que se denomina IA general, atisbo que estamos dispuestos a permitir, sin un debate serio, que, siglos de avances en implantación de derechos humanos, sean ignorados o incluso violentados por el uso de sistemas

---

<sup>51</sup> Gary Marcus. <https://hipertextual.com/2022/06/inteligencia-artificial-conciencia-expertos>

<sup>52</sup> Byung- Chul Han. No-Cosas . Pág. 59.

<sup>53</sup> Citando la lógica de Hegel. Byung- Chul Han. Opus cit. Fig.57.

<sup>54</sup> Markus Gabriel. Ética para tiempos oscuros. Valores universales para el siglo XXI. Pág. 12.



desarrollados sin una regulación que garantice el respeto a aquellos, dejando en manos de su creador atender o no a los límites éticos en su diseño, desarrollo o implementación.

Se advierte una deriva silente hacia un totalitarismo tecnológico provocado por el consumo bulímico digital de una sociedad anestesiada por las últimas aplicaciones que les alteran sutilmente sus aptitudes y actitudes, modificando capacidad de atención, consciencia real y juicio crítico, a modo de la pastilla soma de 1984. Como indica el profesor Lassalle<sup>55</sup>, "estamos ante un legislador discreto que ejerce una soberanía de facto sobre la información que circula por internet, un instrumento de programación que recrea sin cobertura legal la naturaleza humana al hacerla dependiente de un arte de composición técnica que se guarda bajo siete llaves y cuya inaccesibilidad refuerza una relación absolutamente desequilibrada entre la plataforma y el usuario de ello".

Consciente de las repercusiones en la mente humana positivas y negativas profundas y dinámicas de la IA en la sociedad, debido, en parte a las nuevas formas que su utilización influye en el pensamiento, a las interacciones y la adopción de decisiones de los seres humanos, afectando a la educación, las ciencias sociales y humanas, las ciencias exactas y naturales, la cultura y la comunicación y la información, la UNESCO elaboró una recomendación sobre la ética de la IA, en noviembre de 2021, con el fin de estimular una investigación y una innovación realizadas de manera ética que afiancen las tecnologías de la IA en los derechos humanos y las libertades fundamentales, los valores, los principios y la reflexión moral y ética<sup>56</sup>.

La Resolución del Parlamento Europeo, de 3 de mayo de 2022<sup>57</sup>, sobre la IA en la era digital, advirtió sobre la utilización por regímenes autoritarios de sistemas de IA para controlar, espiar, seguir y clasificar a sus ciudadanos, restringir su libertad de circulación y ejercer vigilancia colectiva. Recuerda que cualquier forma de marcaje normativo ciudadano por parte de las autoridades públicas, especialmente en el ámbito de las fuerzas de seguridad del Estado, el control de las fronteras y el poder judicial, y su uso por parte de empresas privadas o particulares, conduce a la pérdida de autonomía y privacidad y no está en consonancia con los valores europeos<sup>58</sup>. Reclama de los legisladores abordar los riesgos que plantea actualmente la toma de decisiones basada en la IA, ya que ha quedado claramente patente la existencia de efectos nocivos, como la

---

<sup>55</sup>Lassalle Ruiz José María, Humanismo tecnológico y ciberleviatán. Una respuesta cívica a la distopía digital. El derecho y la inteligencia artificial. Cuadernos Digitales de Formación.

<sup>56</sup> <https://www.unesco.org/es/artificial-intelligence/recommendation-ethics>

<sup>57</sup> [https://www.europarl.europa.eu/doceo/document/TA-9-2022-0140\\_ES.html](https://www.europarl.europa.eu/doceo/document/TA-9-2022-0140_ES.html)

<sup>58</sup> Identifica como valores fundamentales europeos, los principios de transparencia, explicabilidad, equidad, rendición de cuentas, responsabilidad y fiabilidad, así como el principio de que la IA y la robótica deben centrarse en el ser humano y desarrollarse para complementar a los seres humanos

discriminación racial y sexual, que son atribuibles a casos concretos en los que la IA se ha implantado sin salvaguardas.<sup>59</sup>

En definitiva, y dado que es imposible plantear un futuro inmediato sin que los sistemas de IA hayan invadido todas las esferas de la vida pública, atendidos los proyectos en los que se pretende sea efectivo su empleo en Justicia, como juristas es nuestra obligación ética procurar que la IA se ajuste a unos principios éticos que garanticen una protección adecuada de las libertades públicas, incluidas la libertad de expresión, la intimidad y que no discrimine y sea respetuosa con los derechos fundamentales.

La Carta ética europea sobre el uso de la inteligencia artificial en los sistemas judiciales y su entorno, identifica cinco principios a los que ha de ajustarse los sistemas de IA:

PRINCIPIO DE RESPETO POR LOS DERECHOS FUNDAMENTALES, garantizando que el diseño y la implementación de herramientas y servicios de inteligencia artificial sean compatibles con aquellos.

PRINCIPIO DE NO DISCRIMINACIÓN, previniendo específicamente el desarrollo o intensificación de cualquier discriminación entre individuos o grupos de individuos.

PRINCIPIO DE CALIDAD Y SEGURIDAD, respecto al procesamiento de decisiones y datos judiciales, utilice fuentes certificadas y datos intangibles con modelos elaborados de manera multidisciplinaria, en un entorno tecnológico seguro.

PRINCIPIO DE TRANSPARENCIA, IMPARCIALIDAD Y JUSTICIA, haciendo que los métodos de procesamiento de datos sean accesibles y comprensibles, autorizando auditorías externas.

PRINCIPIO" BAJO CONTROL DEL USUARIO": excluyendo un enfoque prescriptivo y garantizando que los usuarios sean actores informados y que controlen las elecciones realizadas.

Recordemos que en la actualidad no se encuentra regulada y la propia Administración, como se ha indicado, prevé su implementación a corto plazo, y ante ello ha de convenirse que es exigible a los poderes públicos que ese despliegue tecnológico se ha de efectuar conservando los controles y garantías de la actividad administrativa adaptadas a la nueva realidad, sin olvidar que, el uso indebido de la IA en la aplicación de la ley puede dar lugar a una discriminación automatizada y a un trato ilegal de los ciudadanos, que incrementa el riesgo sistémico y el desequilibrio de poder que podrían afectar a la democracia<sup>60</sup>.

---

<sup>59</sup> Refiere también, que tecnologías como la cibervigilancia y el reconocimiento biométrico, que pueden utilizarse para estos fines, han sido vendidas por empresas de la Unión a regímenes autoritarios de países no pertenecientes a la misma.

<sup>60</sup> Eurojust y la agencia de los derechos humanos europea están impulsando un conjunto de herramientas basadas en principios universales de responsabilidad para el uso de la IA por parte de los profesionales de la justicia capacitarlos para que comprendan mejor las oportunidades, los riesgos y las repercusiones sociales, jurídicas y éticas de la IA

Al reclamar a la Comisión y a los Estados miembros que garanticen la protección adecuada de los derechos y el bienestar de los trabajadores, como la no discriminación, la intimidad, la autonomía y la dignidad humana en el uso de la IA y en la gestión algorítmica, también con respecto a las prácticas de vigilancia indebida, la transparencia en su uso, los algoritmos siempre deben estar sujetos a la supervisión humana, evitando con ello una aplicación positiva de la norma descontextualizada de cualquier elemento ético que le otorga la conciencia humana, añadiendo que las decisiones deben ser responsables, impugnables y, si procede, reversibles, lo que conlleva necesariamente la formación de los desarrolladores de algoritmos en cuestiones relativas a la ética, la transparencia y la no discriminación<sup>61</sup>.

Y en definitiva, si se integra en la función jurisdiccional, se ha de reconocer un nuevo derecho procesal, el derecho a un proceso tecnológico transparente y fiable.

### **Conclusión.**

Los sistemas de IA nos ofrecen una oportunidad inmejorable de mejorar la eficacia de la Justicia, siempre y cuando cumplan con los principios éticos indicados, respeten los derechos fundamentales y no atenten contra la independencia del poder judicial.

Como indicó la CEPEJ: Un sistema judicial acorde con su tiempo sería uno capaz de establecer, administrar y garantizar una verdadera ciberética tanto para el sector público como para el privado, e insistiendo en la total transparencia y equidad en el funcionamiento de los algoritmos, que pueden contribuir un día a la toma de decisiones judiciales.

Dicho lo cual:

La toma de decisión ha de ser adoptada y controlada por el ser humano. Solo un juez puede garantizar el respeto genuino por los derechos fundamentales, equilibrar los intereses en conflicto y reflejar los cambios constantes de la sociedad en el análisis de un caso.

Ningún asunto, por simple que pueda parecer o que se pudiese categorizar debido a la escasez de su cuantía o el mínimo margen decisorio, puede ser adoptado por un sistema de IA sin que éste resulte totalmente transparente desde su diseño y pueda ser explicada la propuesta que ofrece. Propuesta que pueda ser además, revisada, modificada o rechazada por el ser humano.

La utilización de sistemas de IA durante el proceso judicial, con la finalidad de agilizar el mismo, requerirán de las mismas exigencias de transparencia y explicabilidad, con posibilidad de revisión y rectificación humanas. Al mismo tiempo es importante que los jueces dicten resoluciones basadas en sistemas de IA, entendiendo

---

<sup>61</sup> [https://www.europarl.europa.eu/doceo/document/TA-9-2022-0140\\_ES.html](https://www.europarl.europa.eu/doceo/document/TA-9-2022-0140_ES.html). Apartado 203

completamente el funcionamiento de la misma, con acceso a toda la información y datos que han sido considerados, para poder asumir la propuesta y explicar sus decisiones.

La ética del funcionamiento del sistema de IA aplicado en Justicia residencia en mejorar la eficiencia y fiabilidad de aquella bajo la supervisión humana, implicando además la obligación de conocer y comprender su funcionamiento.