

Trabajo Fin de Grado

Justicia robótica: ¿herramienta o sustituto?

Autor

Verónica Pardos Gracia

Director

Alberto Lafuente Torralba

Facultad de Derecho

2024

Índice

I. Listado de abreviaturas utilizadas	5
II. Introducción.	5
III. Motivos de la elección del tema a tratar.	5
1. Metodología:.....	6
IV. Definiciones:	6
1. Inteligencia Artificial.	6
2. Algoritmos.....	8
3. Principio de transparencia en el entorno digital	9
4. Sesgos.....	11
5. Paradigma de Inteligencia Artificial Generativa: ChatGPT3.....	11
6. Funcionamiento.	12
7. Posibles Riesgos.	14
8. Responsabilidad en caso de incorrección de la IA	16
IV. Breve introducción a la regulación actual sobre Inteligencia Artificial en la Administración de Justicia.	18
1. Entorno Europeo	18
1.1. <i>Reglamento sobre la Inteligencia Artificial</i>	18
1.2. <i>Libro Blanco sobre la Inteligencia Artificial de la Comisión Europea</i>	19
1.3. <i>Reglamento General de Protección de Datos (en adelante, RGPD)</i>	20
2. Entorno Nacional.....	21
2.1. <i>Constitución Española: ¿Exigencia de un Juez con rostro humano?</i>	21
V. Inteligencia Artificial como sustituto del magistrado.	21
1. Introducción:.....	21
2. ¿Qué es un Juez y cuál es su labor?.....	24
3. Inteligencia Artificial como evaluador y juzgador. Frío, pero imparcial.	28
4. Análisis de caso: Jueces robot en Estonia y PROMETEA en Argentina.	32
4.1. Programa piloto de “jueces robot” en Estonia, especializados en conflictos de baja cuantía y con posibilidad de recurrir la decisión ante un tribunal humano.	32
4.2. PROMETEA, análisis aplicado al caso del uso de la IA como asistente del Sector Público.	33
VI. Inteligencia Artificial como asistente del magistrado.....	37
1. Introducción.....	37
2. Asistencia en la redacción de fallos y resumen de la información. Reino Unido a la cabeza y una España tímida frente al avance.	38
3. Análisis de jurisprudencia en el entorno español: Posición mayoritaria.	40

4. Análisis de caso: COMPAS en Estados Unidos y Viogén en España.	41
4.1. COMPAS en el Caso Loomis, análisis aplicado al caso de la desigualdad y de los sesgos algorítmicos en el uso de la IA.	41
4.2. VIOGEN, análisis aplicado al caso del uso de la IA como predictor de reincidencia en casos de Violencia de Género.	47
VII. Inteligencia Artificial como asistente del abogado.	49
1. Mejora de eficiencia: asistente jurídico.	49
2. Otros asistentes jurídicos existentes hoy en día, destinados a facilitar las tareas del abogado.	54
VIII. Conclusiones.....	55
IX. Anexo I.....	62
X. Bibliografía	66

"Me gustan los robots. Me gustan mucho más que los seres humanos. Si fuese posible crear un robot capaz de ser funcionario civil, creo que haríamos un gran bien, ya que las Leyes de la Robótica le impedirían dañar a un ser humano, lo incapacitarían para la tiranía, la corrupción, la estupidez, el prejuicio".

"La evidencia", Isaac Asimov (Yo Robot)

A mi padre, a mis hermanas y a mi pareja, gracias a quienes soy quien soy y hacia quienes sólo puedo expresar mi sincero agradecimiento por apoyarme durante la etapa académica que hoy culmina.

I. Listado de abreviaturas utilizadas

IA: Inteligencia Artificial.

UE: Unión Europea.

RIA: Reglamento de Inteligencia Artificial, de 13 de marzo de 2024.

NLP (inglés)/ PLN (español): Procesamiento de Lenguaje Natural.

Q&A: Preguntas y respuestas.

TSJ: Tribunal Superior de Justicia.

II. Introducción.

El papel central de los jueces y magistrados en nuestro sistema jurídico es innegable. Sin embargo, en los últimos tiempos, este sistema clásico, que ha perdurado durante décadas, enfrenta amenazas sin precedentes. Una pandemia mundial nos obligó a adaptarnos rápidamente a una justicia telemática, mientras que la sobrecarga y las limitaciones de tiempo de los jueces han generado listas de espera interminables, dejando a familias en incertidumbre durante años.

Este paradigma, que hasta ahora había funcionado sin apenas cuestionamientos, se encuentra ahora al borde de una transformación radical impulsada por el avance tecnológico: la Inteligencia Artificial.

El sistema jurídico es una de las bases fundamentales de cualquier sociedad y, como tal, cualquier cambio en este ámbito debe ser tratado con la máxima cautela. En este Trabajo Final de Grado exploraré los nuevos límites de esta realidad y las posibilidades de implementar esta tecnología tanto a nivel nacional como internacional. Examinaré los dilemas éticos que plantea, así como los retos y beneficios de su adopción. Todo esto se analizará desde varias perspectivas: el uso de la Inteligencia Artificial como asistente del magistrado, manteniendo la justicia en manos humanas; y la posibilidad de que la Inteligencia Artificial reemplace al juez como elemento central de nuestro sistema, generando un paradigma para el que, según la opinión de muchos, todavía no estamos preparados. Por último, se realizará un somero análisis sobre las posibilidades que se presentan a los abogados para implementar la Inteligencia Artificial como una innovadora herramienta en sus funciones.

III. Motivos de la elección del tema a tratar.

El objetivo de este trabajo es acercar al lector las posibilidades que la tecnología ofrece a los diversos profesionales del derecho, con especial atención a como esta afecta a jueces, magistrados, y abogados. A la vez, pretende plantear los retos que este cambio de paradigma supone en el presente y aquellos que se prevé que genere en el futuro. Este tema, de máxima actualidad, aún permanece fuera del foco público, quizás por miedo o cautela, pero sin duda formará parte de nuestras vidas dentro de unos años.

Con estas páginas, se busca concienciar, o al menos divulgar, un futuro que ya está aquí.

1. Metodología:

Para la escritura de estas páginas se ha realizado una extensa recopilación documental, lectura e interiorización de un tema tan complejo y abierto a opiniones diversas, tras lo que se ha comenzado con la escritura de un primer borrador de los epígrafes.

Poco a poco, conforme la comprensión era más profunda y las reflexiones, de mayor calidad, se ha tratado de plasmar con precisión y exactitud los temas más relevantes en el desarrollo de la IA en la Administración de Justicia, de manera objetiva en un inicio, pero añadiendo una reflexión personal cuando se ha considerado necesario. La parte objetiva y subjetiva quedan siempre diferenciadas, con la intención de que el lector pueda tanto informarse, como adentrarse en mis reflexiones.

IV. Definiciones:

1. Inteligencia Artificial.

Mucho se ha hablado en los últimos tiempos sobre la Inteligencia Artificial (en adelante, IA). Este concepto, en boca de todos y relacionado, al parecer, con todos los problemas y los avances de esta década es, paradójicamente, todavía un perfecto desconocido.

Y ello se debe a que, al igual que la inteligencia humana, la Inteligencia Artificial es un concepto dinámico y difícil de acotar.

La Comisión Europea la define en el Libro Blanco sobre la Inteligencia Artificial¹ como aquellos “sistemas de software (y posiblemente también de hardware) diseñados por humanos que, ante un objetivo complejo, actúan en la dimensión física o digital:

- Percibiendo su entorno, a partir de la adquisición e interpretación de datos estructurados o no estructurados
- Razonando sobre el conocimiento, procesando la información derivada de estos datos y decidiendo las mejores acciones para lograr el objetivo dado.

En un sentido similar, el recién aprobado Reglamento de la IA² ha definido, por primera vez, los “Sistemas de Inteligencia Artificial” desde una perspectiva jurídica, indicando que estos son aquellos sistemas basados en una máquina diseñados para funcionar con distintos niveles de autonomía, que pueden mostrar capacidad de adaptación tras el despliegue y que, para objetivos explícitos o implícitos, infieren de la información de entrada que reciben la manera de generar información de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales. Además, los Sistemas de IA pueden usar reglas simbólicas o aprender un modelo numérico. Dicho de otro modo: la IA es un campo de la informática que se enfoca en crear sistemas que puedan realizar tareas que normalmente requieren inteligencia humana, como el aprendizaje, el razonamiento y la percepción, pudiendo estos sistemas percibir su entorno, razonar sobre el conocimiento, procesar la información derivada de

¹ Definición consensuada por un grupo de expertos, que la incluyeron por primera vez en el documento de Directrices éticas para una IA fiable. Es, sin embargo, una cuestión controvertida la generación de una definición omnicomprensiva y útil para referenciar el cambiante e ingente sistema que comprende la “Inteligencia Artificial”.

² Aprobado definitivamente el 13 de marzo de 2024.

los datos y tomar decisiones para lograr un objetivo dado. Sin embargo este término, ya acuñado por John McCarthy en 1956 durante la Conferencia de Dartmouth, avanza y evoluciona con rapidez, por lo que las clasificaciones de este varían en consonancia, mutables y abiertas a las nuevas invenciones.

Aquella clasificación que nos interesa, a efectos de este trabajo, es aquella que diferencia la IA en Débil, Fuerte y Superinteligente, con especial foco en la segunda de ellas.

La **IA Débil** es capaz de realizar únicamente tareas específicas y limitadas, como la traducción entre idiomas. No tiene capacidad de aprendizaje o adaptación por sí misma, y necesita ser programada para realizar una tarea determinada. Su alcance es, por ello, limitado a la tarea para la que se programó.

La **IA Fuerte** está diseñada para tener capacidad de aprendizaje autónomo, siendo capaz de realizar múltiples tareas y aprender de manera autónoma a medida que interactúa con su entorno. Es lo que vulgarmente se denomina como “aprender de sí misma”, es decir, mejorar la calidad de sus predicciones y adaptarse a la tarea dada mediante la realización de esta y la recepción de *feedback* por los programadores, hasta conseguir identificar con certeza las claves para la resolución de sus casos, generando soluciones potencialmente óptimas por sí misma.

Por último, la denominada **IA Superinteligente**³ es un tipo de IA que superaría la inteligencia humana en muchos (o todos) los aspectos. Esta Inteligencia Artificial sería capaz de comprender el mundo de una manera diversa a la humana, pudiendo resolver problemas complejos a una velocidad y eficiencia superiores a los seres humanos. Este tipo de IA es todavía una forma teórica, no desarrollada en la práctica, pero es aquella que se pretende utilizar en el caso de pretender que esta “sustituya” a los humanos en operaciones que requieren, comúnmente, de inteligencia social.

Ejemplo de esto sería el caso de que se pretenda que la IA sustituya a un juez. Como se expondrá en el apartado correspondiente, la función de un juez es compleja y difícilmente automatizable, debido al necesario margen de discrecionalidad que se precisa para la interpretación y aplicación de las leyes. Mientras que esto podría suponer un reto para una IA fuerte, una IA Superinteligente sería capaz de mimetizar el razonamiento detrás de cada interpretación, desarrollando la capacidad de “pensar” (opuesta a la aplicación de criterios cerrados e inmutables, como los requeridos para realizar una operación matemática) y obtener una solución realmente similar a lo que un ser humano habría conseguido, pero en menos tiempo e, idealmente, con menos sesgos (tema de vital importancia que también será tratado en el apartado correspondiente).

Como se puede apreciar, no son pocos los dilemas éticos, morales y sociales que el desarrollo e implementación de esta IA genera, lo que, sumado a las limitaciones técnicas que derivan de la novedad de las IA y de la complejidad de este nuevo paradigma, genera dudas incluso sobre la posibilidad de que, en el futuro, seamos capaces de desarrollar este modelo de IA, que por ahora reside únicamente en el mundo de lo teórico.

³ Definición realizada, entre otros en el mismo sentido, por BARONA VILAR, S., en *La seductora algoritmización de la justicia. Hacia una justicia poshumanista (Justicia+) ¿utópica o distópica?*, El Cronista del Estado Social y Democrático de Derecho, Nº 100, 2022 pág 36 a 46.

2. Algoritmos.⁴

Toda IA se fundamenta en dos pilares necesarios: los datos y los algoritmos. Los datos son el conjunto de registros de casos anteriores que los programadores deben dar a la IA, para que esta los analice y los utilice como resultado a obtener en sus tareas (por ejemplo, si se pretende entrenar a una IA para que resuma textos, se le debe proveer del texto en crudo y del resumen tal como lo haría un humano, para que “aprenda” cómo debe realizar su tarea de una forma óptima). Los algoritmos son las “instrucciones” que indican a la máquina qué hacer. Existen varios tipos de algoritmo que debemos conocer antes de tratar cómo estos pueden influir en la administración de Justicia.

- **Reinforcement Learning** (Aprendizaje por refuerzo).

Este tipo de algoritmo indica a la máquina que debe aprender mediante el sistema de “prueba y error”. La máquina puede realizar, a velocidades increíbles, múltiples pruebas y aprender de sus resultados. Para ello, se la programará con las condiciones del “juego” y se establecerá cuál es el objetivo por alcanzar, denominado “recompensa”. Así, la máquina realizará un gran número de veces el “juego”, de acuerdo con las reglas establecidas, generando patrones, correlaciones, caminos y conclusiones basadas en las experiencias generadas por la propia máquina. Un ejemplo puntero de este tipo de aprendizaje lo encontramos en la IA ajedrecista *AlphaZero*, de *DeepMind*, especialista en ganar partidas de ajedrez.

- **Supervised machine learning** (Aprendizaje supervisado).⁵

En este caso, se utilizan modelos predictivos a partir de datos de entrenamiento. Es decir, se introducen un conjunto de datos etiquetados o clasificados para entrenar al algoritmo a realizar tareas concretas. Así, los algoritmos aprenden de datos históricos, sin necesidad de probar todas las opciones posibles sino infiriendo patrones de los datos aportados. Este modelo se utiliza en el aprendizaje de los coches automáticos.

- **Unsupervised machine learning** (Aprendizaje no supervisado).⁶

Este tipo de algoritmo se utiliza para identificar nuevos patrones y anomalías. Los datos introducidos en la IA no están etiquetados, pues se busca que sea el propio algoritmo quien intente dar sentido a los datos a través de la búsqueda de patrones y características. Este tipo de procesamiento de datos es utilizado para la gestión de grandes cantidades de datos, que la IA será capaz de analizar desde diversas perspectivas para tratar de encontrar patrones que no habían sido identificados hasta el momento. Usos comunes de este tipo de aprendizaje son, entre otros, la detección del fraude, la segmentación entre diferentes tipos de clientes o el estudio genético.

⁴ Google es una de las compañías que más interés ha tenido desde lejano en tratar de explicar al público qué son los algoritmos, en aras de cumplir con el principio de transparencia al que se ve obligado por ley. En una serie de artículos científicos de “*GoogleCloud*” el lector interesado puede encontrar una gran cantidad de referencias a los diferentes tipos existentes de algoritmos, sus usos y sus riesgos, en la página web Cloud Google, apartado *Discover*.

⁵ *Supervised vs. Unsupervised learning: What’s the difference?* Google Cloud, 2024 y *What is Supervised Learning?* Google Cloud, 2024.

⁶ *What is unsupervised learning?* Google Cloud, 2024.

3. Principio de transparencia en el entorno digital.

La transparencia es un principio fundamental y, a menudo, uno de los más complejos de respetar en el entorno digital. Con el avance de la Inteligencia Artificial, han surgido desafíos significativos para garantizar una protección efectiva de este principio pues, desde el inicio, la transparencia ha sido reconocida como un derecho esencial del usuario y es de obligatorio cumplimiento según el Reglamento de Inteligencia Artificial de la UE⁷.

En concreto, la transparencia en el ámbito de la Inteligencia Artificial se refiere a la información que deben suministrar los proveedores de sistemas de Inteligencia Artificial a sus usuarios, entendiéndose como tales las entidades y entes que van a utilizar dicho sistema de IA. Es decir, este principio pretende que los desarrolladores de las IA proporcionen información suficiente relativa a sus características, funcionamiento y capacidades. Este requerimiento plantea diversos problemas, siendo uno de los principales el derecho a la propiedad industrial, pues si la obligación de revelación del funcionamiento de los sistemas fuera absoluta, ninguna empresa invertiría tanto dinero como se requiere en desarrollo de software, sino que esperaría a que otra empresa realizara la inversión para después, cuando la información sobre el funcionamiento de las novedades fuera pública, copiar el procedimiento a seguir.⁸

Por ello, este principio de necesidad de transparencia no es absoluto. De acuerdo con el Reglamento de la IA, la transparencia implica que “los sistemas de IA se desarrollarán y utilizarán de forma que permitan una trazabilidad y explicabilidad adecuadas, al tiempo que se hace saber a los seres humanos que se comunican o interactúan con un sistema de IA y se informa debidamente a los usuarios de las capacidades y limitaciones de dicho sistema de IA y a las personas afectadas de sus derechos”.

La cuestión de tratar de discernir hasta qué punto ha de ponderar este derecho del usuario ha quedado regulada por el Reglamento de la IA de la UE, en el que se establecen las principales obligaciones respecto a la transparencia en la IA:

- Los sistemas de IA de alto riesgo⁹ deben diseñarse y desarrollarse de manera que garanticen un funcionamiento con un nivel de transparencia suficiente para que los usuarios interpreten y usen correctamente su información de salida¹⁰.
- Los sistemas de IA de alto riesgo deben ir acompañados de las instrucciones de uso correspondientes, que incluirán la información concisa, completa, correcta y

⁷ Resolución legislativa del Parlamento Europeo, de 13 de marzo de 2024, sobre la 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. Exposición de motivos, Capítulo IV, “Obligaciones de transparencia de los proveedores y responsables del despliegue de determinados sistemas de IA”.

⁸ Al efecto, el esclarecedor trabajo de NAVAS NAVARRO, S, *Responsabilidad civil e Inteligencia Artificial*, el Cronista del Estado Social y Democrático de Derecho N° 100, págs. 196 a 110; y NAVAS NAVARRO, S, *Salud e Inteligencia artificial desde el Derecho privado*, Comares, Granada, 2021.

⁹ Como se explicará en el apartado correspondiente, a efectos del Reglamento de la IA, se consideran sistemas IA de alto riesgo aquellos utilizados en infraestructuras críticas, formación educativa, componentes de productos, sistemas de conducción autónoma y aquellos utilizados por los bancos para evaluar la solvencia y la calificación crediticia de los ciudadanos.

¹⁰ Considerando 73 Reglamento IA

clara que sea pertinente, accesible y comprensible para los usuarios. Esta información especificará¹¹:

- La identidad y los datos de contacto del proveedor.
 - Las características, capacidades y limitaciones del funcionamiento del sistema de IA, en concreto:
 - Su finalidad prevista.
 - El nivel de precisión, solidez y ciberseguridad.
 - Cualquier circunstancia conocida o previsible asociada al uso del sistema de IA conforme a su finalidad prevista o a un uso indebido razonablemente previsible, que pueda dar lugar a riesgos para la salud y la seguridad o los derechos fundamentales.
 - Su funcionamiento en relación con las personas o los grupos de personas en relación con los que se pretenda usar el sistema.
 - Cuando proceda, especificaciones relativas a los datos de entrada o cualquier otra información pertinente respecto a los conjuntos de datos de entrenamiento, validación y prueba usados.
- Los proveedores de sistemas de IA destinados a interactuar con personas físicas deben garantizar que dichas personas estarán informadas de que están interactuando con un sistema de IA (salvo por las circunstancias y el contexto, resulte evidente).¹²
 - Los usuarios de un sistema de reconocimiento de emociones o de un sistema de categorización biométrica informarán del funcionamiento del mismo a las personas físicas expuestas a él.¹³
 - Los usuarios que empleen sistemas de IA para manipular o generar contenido de imagen, sonido o vídeo que se asemeje notablemente a personas, objetos, lugares u otras entidades o sucesos existentes y que puedan conducir a una persona a pensar que son auténticos o verídicos, deben hacer público que el contenido ha sido generado o manipulado con IA.¹⁴

Así pues, y frente al usuario final que utiliza una IA, la información que habrá de estar siempre disponible será: indicar cuando estos están ante un sistema de IA, indicar los efectos que puede tener su uso respecto a las decisiones automatizadas (falta de precisión, falta de veracidad), e informar a estos cuando estén expuestos a material audiovisual creado por IA.

La información técnica, relativa a las instrucciones de funcionamiento concreto del sistema de IA, los registros y la gestión de los datos de entrada y salida y los riesgos para la seguridad, salud y los derechos fundamentales de los usuarios finales, estará destinada a las entidades que desplegarán los sistemas de IA y no a los usuarios finales¹⁵.

¹¹ Considerando 72 Reglamento IA

¹² Art 50.1 Reglamento IA

¹³ Art 50.3 Reglamento IA

¹⁴ Art 50.2 Reglamento IA

¹⁵ Siendo estas entidades aquellas empresas (públicas o privadas) que han contratado la creación de un servicio de IA que van a poner a disposición de su público.

4. Sesgos

Los sesgos son un peligroso enemigo de la Inteligencia Artificial, así como lo son de las personas en general. Por el propio funcionamiento de la IA (que se nutre de datos de casos anteriores resueltos por humanos, identifica patrones y los utiliza en sus resoluciones posteriores) esta es extremadamente sensible a tener sesgos, que distorsionan los resultados obtenidos¹⁶.

El caso más habitual se produce cuando la máquina se alimenta de datos sesgados (por ejemplo, de datos en los que se imponen penas más duras a personas de una cierta nacionalidad¹⁷; o en los que la custodia de los niños se otorga en una gran cantidad de ocasiones a la madre, en lugar del padre). La máquina va a analizar estas resoluciones, sin juzgarlas correctas e incorrectas, sino entendiéndolas como datos en los que fundamentar su conducta. Así pues, la IA resolverá sus casos posteriores introduciendo dichos sesgos en sus respuestas, respuestas que pasarán a formar parte de su propia base de datos y que confirmarán y amplificarán dicho sesgo, el cual, con el tiempo, generará una situación potencialmente injusta.

En definitiva, si no se opera con extremo cuidado, es sencillo que los modelos en que se basa una IA absorban los prejuicios de la sociedad, pues estos pueden encontrarse incrustados de manera silenciosa en las montañas de datos con los que las IA se entrenan.¹⁸

5. Paradigma de Inteligencia Artificial Generativa: ChatGPT3.

ChatGPT es un modelo de lenguaje desarrollado por OpenAI basado en la arquitectura GPT-4 (*Generative Pre-trained Transformer 4*). Está diseñado para comprender y generar texto de manera similar a cómo lo haría un ser humano, utilizando técnicas avanzadas de aprendizaje automático y procesamiento de lenguaje natural (PNL)¹⁹. Este

¹⁶ Como con acierto apunta NAVAS NAVARRO, S, *Daños ocasionados por sistemas de inteligencia artificial. Especial atención a su futura regulación*, Comares, Granada, 2022 y RIVAS VELASCO, M.J, *Uso ético de la Inteligencia Artificial en Justicia*, Diario La Ley, Sección Tribuna, págs. 5 a 10, 13 de julio de 2023.

¹⁷ Iturmendi Rubia, J.M. *Algorithmic discrimination and its impact on human dignity and human rights. Special reference to immigrants*, 2023.

¹⁸ <https://www.ibm.com/topics/aibias#:~:text=El%20sesgo%20de%20IA%2C%20tambi%C3%A9n,resultados%20distorsionados%20y%20potencialmente%20perjudiciales>.

¹⁹ El PNL es una rama de la IA que se centra en la mejora de la interacción entre seres humanos y máquinas. Su objetivo principal es permitir que las máquinas entiendan, interpreten y respondan al lenguaje humano de una manera útil y natural. Para ello, el PNL somete a la máquina que se pretende entrenar a diferentes análisis, los cuales, poco a poco, ayudan a refinar el entendimiento de la máquina del lenguaje humano, para que los resultados aportados por esta sean, cada vez, más cercanos a los que podrían ser generados por un humano.

Ejemplo de estos análisis son:

- Análisis morfológico, el cual examina la estructura interna de las palabras para entender su composición y formación.
- Análisis sintáctico, enfocándose en la estructura gramatical de las frases y determinando cómo las palabras se combinan entre sí para formar oraciones correctas.

Existen, además, otros análisis más complejos y de reciente introducción, pero que han sido los responsables de que las máquinas estén, poco a poco, cada vez más cerca de superar por completo el Test de Turing. Ejemplo de estos nuevos análisis son:

- Análisis pragmático, capaz de analizar el texto en el cual se utiliza el lenguaje para interpretar significados más allá de las palabras literales.

modelo cuenta con más de 175 millones de parámetros, y ha sido entrenado con ingentes cantidades de texto que le permiten realizar con altas tasas de eficacia diversas tareas relacionadas con el lenguaje²⁰. Además, desde septiembre de 2023 existe una nueva actualización que permite a ChatGPT utilizar información de internet hasta junio de 2023 (en lugar de desde 2021, como ocurría antes de esta actualización).

6. Funcionamiento.

Para que el conocido chat conteste al público tras la realización de sus preguntas, son necesarias diversas etapas de entrenamiento y ajuste. Además, resulta importante destacar que, pese a la gran revolución que esta herramienta ha causado, esta sigue siendo un apoyo que ha de utilizarse, siempre, en consonancia con otras fuentes de datos pues, como se explicará a continuación, las respuestas dadas por esta IA generativa se basan simplemente en un cálculo de probabilidad de la respuesta potencialmente más adecuada para cada caso, sin hacerse en ningún caso responsable la plataforma de las incorrecciones o inexactitudes de las que la IA pueda proveer al usuario.²¹

A continuación se realiza un breve repaso sobre las fases por las que la IA ChatGPT tuvo que pasar antes de ser viable para su utilización por el público, y por las que pasa en cada actualización:

Preentrenamiento

1. Creación de un *Corpus* de Datos:

ChatGPT se entrena inicialmente con una gran cantidad de texto proveniente de diversas fuentes en Internet. Este texto puede incluir libros, artículos, sitios web, foros, entre otros. Simplemente, se da acceso al modelo a esta información, para que la máquina pueda moverse libremente dentro de esta.

2. Entrenamiento No Supervisado:

En esta fase, el modelo se entrena de manera no supervisada. Esto significa que no hay etiquetas específicas que el modelo deba aprender. En lugar de eso, aprende a predecir la siguiente palabra en una oración a partir de prueba y error. Por ejemplo, si el texto es "El gato se subió al", el modelo aprenderá a predecir "árbol" como una posible continuación (al ser esta una respuesta muy repetida en su *corpus* de datos). Mediante este entrenamiento, la máquina aprende a generar un *ránking* sobre cómo decidir qué respuesta

-
- Análisis de sentimientos, al posibilitar la interpretación de emociones o actitudes expresadas en un texto.
 - Análisis discursivo, responsable de examinar como las oraciones se conectan entre sí dentro de un texto para entender el significado de un contexto más amplio.

Antonio Moreno, *Procesamiento del lenguaje natural ¿qué es?*, Instituto de Ingeniería del Conocimiento.

²⁰ Entre las que se encuentran la traducción de texto, la generación de resúmenes, la ampliación de un texto dado por el usuario y cientos de opciones más.

²¹ Desde la última actualización (marzo de 2024), el conocido chat ha añadido, debajo de la barra de preguntas para ChatGPT, un *disclaimer* en el que se indica "ChatGPT puede cometer errores. Considera verificar la información importante".

es mejor para cada pregunta, comparando sus suposiciones con los datos de internet (que posee gracias al punto 1), para, poco a poco, pulir su criterio.

3. **Objetivo de Modelado del Lenguaje (NLP):**

El objetivo del preentrenamiento es que el modelo aprenda las estructuras del lenguaje, como gramática, hechos de conocimiento general y algunas habilidades de razonamiento y abstracción²². Es precisamente gracias a que esta fase tenga tanta importancia que el modelo parezca tan real, y que sus respuestas sean tan similares a las de un humano (pues se ha entrenado leyendo y analizando datos, textos y demás documentos de internet redactados por humanos, para después ser sometido a diferentes “análisis” que miden su habilidad para entender e interpretar el lenguaje de manera avanzada).

Fine-Tuning (Ajuste Fino)

1. **Conjunto de Datos Específico:**

Después del preentrenamiento, el modelo se ajusta finamente usando un conjunto de datos más pequeño y específico. Este conjunto de datos es curado²³ y puede incluir ejemplos de cómo se espera que responda el modelo a ciertas preguntas o situaciones. Con esta fase, se realiza un especial hincapié en las tareas más complejas, o se corrigen fallos que la IA hubiere podido cometer por su cuenta. Se trata de lijar aquellas asperezas del modelo y de reconducir aquellas actividades que, de acuerdo con el experto, deberían ser realizadas o entendidas de otra manera.

En definitiva, durante el ajuste fino, se utiliza retroalimentación humana para guiar al modelo (mediante técnicas de aprendizaje por refuerzo, definido en el apartado I). Los entrenadores humanos pueden corregir respuestas del modelo, lo cual ayuda a alinear sus respuestas con expectativas humanas y reducir la generación de contenido inapropiado o incorrecto.

Generación de Respuestas

1. **Entrada del Usuario:**

- Cuando un usuario escribe una pregunta o una solicitud, esa entrada se convierte en una secuencia de tokens (fragmentos de palabras o caracteres digitales).

2. **Procesamiento de la Entrada:**

- El modelo procesa estos tokens y utiliza su entrenamiento previo para generar una secuencia de tokens de salida que forman la respuesta.

3. **Mecanismo de Atención:**

- Utiliza un mecanismo de atención para enfocarse en diferentes partes de la secuencia de entrada mientras genera cada token de salida. Esto le

²² Aunque, como con maestría pone en cuestión Javier Sánchez Cañizares en su ensayo *¿Qué nos hace humanos ante la Inteligencia Artificial?*, sería una cuestión debatible si la IA puede, por definición, ser capaz de “abstraerse” (al menos en el sentido en el que lo entendemos comúnmente).

²³ Expresión técnica que se refiere al conjunto de verificaciones y acciones realizadas sobre los datos por personas capacitadas con el fin de asegurar que el conjunto de estos está estructurado y documentado de la manera más completa posible y de acuerdo con los estándares propuestos.

permite manejar contextos largos y recordar detalles relevantes a lo largo de la generación de la respuesta.

4. **Probabilidades de las Palabras:**

- Para cada token de salida, el modelo calcula una probabilidad para cada palabra posible en el vocabulario y selecciona la palabra con la mayor probabilidad de ser la respuesta correcta en cada situación, o realiza una selección basada en una distribución de probabilidad para generar respuestas más variadas y naturales. Para ello, analiza los resultados que ha tenido en casos anteriores, la información dada (de internet, principalmente) y el ajuste fino, para decidir qué siguiente palabra tiene mayor probabilidad de ser la palabra correcta en la frase.

De manera visual, podría verse como una nube de puntos de respuestas potenciales, aportándose una probabilidad a cada respuesta de acuerdo con el entrenamiento de la máquina, que en milésimas de segundo realizará el cálculo de las respuestas potenciales, las ordenará de acuerdo con su criterio, y generará tokens de salida para generar una respuesta a la consulta del usuario.

7. Posibles Riesgos.

ChatGPT se erige, desde luego, como una herramienta útil. Pero la novedad de su funcionamiento puede crear riesgos que los usuarios debemos conocer y contrarrestar mediante un uso responsable de la plataforma. Algunos de los riesgos de mayor impacto potencial se describen a continuación.

Desinformación:

- ChatGPT puede generar respuestas que suenen plausibles pero que sean incorrectas o engañosas. Dado que no tiene acceso a información actualizada en tiempo real (tan solo a la información que existía en internet hasta junio de 2023), su conocimiento está limitado a los datos con los que fue entrenado, lo que puede incluir información desactualizada o directamente incorrecta. Además, internet es público, y ChatGPT recabará información tanto de fuentes fiables como de fuentes no fiables. Si existe información incorrecta, pero muy repetida en internet, ChatGPT puede tomarla como información certera y utilizarla en sus respuestas.

Uso Malintencionado:

- La tecnología puede ser utilizada para generar contenido malicioso, como propaganda, spam, fraudes o desinformación a gran escala. Esto puede amplificar el alcance de campañas de desinformación y afectar negativamente a individuos y sociedades²⁴.

²⁴ Jiménez I. *La era de las estafas inteligentes: Cómo la IA está cambiando el juego del engaño*. <https://www.abogacia.es/publicaciones/blogs/blog-de-innovacion-legal/la-era-de-las-estafas-inteligentes-como-la-ia-esta-cambiando-el-juego-del-engano/>

Privacidad y Seguridad:

- Si no se maneja adecuadamente, ChatGPT puede divulgar información sensible o privada. Aunque no tiene la capacidad de recordar interacciones pasadas más allá de la sesión actual, siempre existe el riesgo de que se generen respuestas que inadvertidamente incluyan datos privados, al haber sido estos aportados por usuarios en sesiones diversas pero relativas a un tema similar.²⁵

Sesgos:

- Al ser entrenado con datos de internet, el modelo puede reflejar los sesgos presentes en esos datos. Esto puede incluir sesgos de género, raciales, culturales y otros, lo que podría llevar a la generación de contenido que perpetúe estereotipos o discriminación.

Dependencia Tecnológica:

- La facilidad de uso y la capacidad de generar texto automáticamente pueden llevar a una dependencia excesiva de esta tecnología, reduciendo la capacidad de pensamiento crítico y la producción de contenido original por parte de los usuarios.

Cabe destacar que, aunque esta herramienta puede ser revolucionaria, no debe ser entendida como un oráculo al que escuchar sin vacilación. Algunos de los problemas que esta IA presenta no son, en ocasiones, suficientemente conocidos por el usuario, que puede hacer un mal uso de la misma al no tenerlos en cuenta:

1. Falta de actualización, por lo que sus respuestas tan sólo estarán informadas hasta junio de 2023. Sin embargo, en ningún momento avisará de este hecho, sino que dará una respuesta por completa, cuando realmente será carente de casi un año de avance en internet.
2. Alucinaciones, que son, simplemente, huecos en el conocimiento de la IA, que esta llenará con información incorrecta. Este fenómeno se debe a que la IA está programada para, siempre que sea posible, dar una respuesta coherente y similar a la que daría un humano, eligiendo los argumentos a utilizar con base probabilística sobre qué resulta más plausible que sea cierto en cada momento. En ciertas ocasiones, si los datos de internet son contradictorios, confusos o simplemente, si no es capaz de encontrarlos, genera una respuesta que, aunque parece plausible, es factualmente incorrecta. De nuevo, ChatGPT no avisa de

²⁵ Como ejemplo, a inicios de 2023, cuando este potencial problema era poco conocido, empleados de Samsung generaron un gran problema de privacidad al introducir datos confidenciales en el chat, pidiéndole al mismo ayuda con la resolución de unos problemas de código. Tras la consulta, dichos datos pasaron a formar parte de las fuentes de información del chat, que los utilizó como parte de la respuesta a las consultas de otros usuarios.

Tras este incidente, empresas como Apple o el propio Samsung han bloqueado el acceso a herramientas de IA a sus empleados.

<https://www.20minutos.es/tecnologia/ciberseguridad/empleados-samsung-filtran-informacion-confidencial-usar-chatgpt-5116256/>

cuándo se están produciendo dichas alucinaciones, sino que las muestra al usuario como si se tratara de datos contrastados²⁶.

8. Responsabilidad en caso de incorrección de la IA

Uno de los dilemas fundamentales y de más compleja resolución al respecto del avance de la IA es el relativo a la responsabilidad por daños en caso de que una IA cause perjuicios, pues estas, a disposición de todo aquel que quiera utilizarlas, no se encuentran reguladas de manera específica en nuestro entorno jurídico. La sociedad, una vez más, ha superado al derecho, y este sin duda se enfrenta a un gran reto: crear una regulación para una realidad que presenta características sin precedentes.

Existe un principio general en nuestro derecho (art. 1902 del Código Civil) que indica que *“el que por acción u omisión causa un daño a otro, interviniendo culpa o negligencia, está obligado a reparar el daño causado”*. Es decir, en términos generales, quien causa un daño es responsable de repararlo. Sin embargo, no existe normativa alguna que adecue este precepto general al ámbito de la Inteligencia Artificial.

Se trata sin duda de un problema jurídico complejo, incluso, respecto a quién sería el causante del daño y quién el responsable según la ley actual ¿Sería el causante del daño el desarrollador del sistema? ¿el propietario de la plataforma? ¿la propia IA?. Pese a que no existe todavía una regulación vinculante, existe una Resolución del Parlamento Europeo de octubre de 2020²⁷, que pretende establecer una normativa respecto a la responsabilidad civil cuando un sistema de IA genere algún daño. Esta Resolución genera recomendaciones destinadas a la Comisión sobre una potencial regulación y presenta, sin duda, ideas interesantes, como una dualidad en el régimen de responsabilidad en función del sistema de IA que ha causado el daño (por un lado los sistemas de IA de alto riesgo²⁸, para los que se establece un sistema de responsabilidad objetiva y, por otro lado, una posible responsabilidad subjetiva para los daños causados con sistemas de IA que no supongan un alto riesgo) o una posible aproximación al concepto de sujeto responsable, denominado “operador”. Sin embargo, estas son tan solo recomendaciones que, desde 2020, no han sido llevadas a la práctica y que parecen estar parcialmente superadas.²⁹

²⁶ En el entorno de la Justicia, ya se ha dado el caso de dos abogados (uno de ellos con más de 30 años de experiencia en el ejercicio de la abogacía) que presentaron un escrito judicial generado utilizando ChatGPT, el cual se inventó una serie de precedentes judiciales inexistentes, que los abogados tomaron como ciertos. Ambos fueron multados por una cuantía de 5.000 dólares al apreciar el judicial que habían ofrecido declaraciones engañosas al tribunal, al haber “obviado conscientemente” las señales que apuntaban a que los casos que habían incluido ChatGPT eran falsos. En la sentencia se subraya que “los abogados hicieron dejación de ciertas responsabilidades cuando presentaron las opiniones judiciales no existentes acompañadas de citas falsas creadas por la herramienta artificial de ChatGPT”. Del mismo modo, en nuestro entorno, la IA generará una serie de sentencias, incluyendo citas de las mismas, que respalden el caso planteado por el sujeto. Sin embargo, la IA no se hace responsable ni garantiza que dichas sentencias sean reales.

²⁷ Resolución del Parlamento Europeo, de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un régimen de responsabilidad civil en materia de inteligencia artificial (2020/2014).

²⁸ Categorización realizada en este Reglamento y en la que se ahonda en el Capítulo V “Breve introducción a la regulación actual sobre Inteligencia Artificial en la Administración de Justicia”.

²⁹ Al efecto, los *breves apuntes a la Propuesta de Reglamento del Parlamento Europeo sobre responsabilidad civil en materia de Inteligencia Artificial*, de Alejandro Zornoza Somolinos.

Otro de los grandes retos que habrá de ser regulado es que, pese a la polémica al efecto que desarrollaré con posterioridad, hoy en día las IA no tienen personalidad jurídica propia. Por ello estas no podrán ser directamente responsables, sino que habrán de serlo en su lugar las empresas de desarrollo, sus fabricantes o, incluso, las personas que han hecho un uso negligente de las IA. Sin embargo, ¿deberían empezar a ser personas jurídicas, con derechos y deberes?

Nadie duda de la relevancia social que en los próximos años estas tecnologías van a tomar en nuestras vidas, pero... ¿cuál va a ser su lugar en nuestro sistema jurídico?

Por último, es necesario mencionar que existe una Propuesta de Directiva sobre la Responsabilidad derivada del uso de la IA³⁰, que propone instaurar reglas en los procedimientos de reclamación de daños y perjuicios, así como la posibilidad de instaurar un seguro obligatorio para el uso de ciertos sistemas de IA³¹ (aunque esta medida parece estar enfocada a casos, por ejemplo, como los accidentes de tráfico, cuando un coche autónomo haya sido el causante de los daños).

Sin embargo, y hasta que este tema sea definitivamente regulado, la incertidumbre es mayúscula. Al no existir regulación, nadie quiere hacerse cargo de la responsabilidad que estos problemáticos supuestos conllevan, lo que supone, como explicaré con posterioridad, la necesidad de ponderar de acuerdo con el caso concreto. Esta técnica ya se está viendo completamente insuficiente, por lo que los propietarios de las IAs están, en ocasiones, tratando de salvarse las espaldas mediante indicaciones como el novedoso *disclaimer* de ChatGPT “ChatGPT puede cometer errores. Considera verificar la información importante”. Si estos, en el momento en que se produzca algún litigio relevante, serán suficientes, será algo que el propio Juez habrá de decidir de acuerdo con su buen saber y entender, hasta que se regule de manera específica el problema de la responsabilidad por incorrección de la Inteligencia Artificial.

³⁰ Propuesta de Directiva del Parlamento Europeo y del Consejo relativa a la adaptación de las normas de responsabilidad civil extracontractual a la Inteligencia Artificial (Directiva sobre responsabilidad en materia de IA).

³¹ A partir del estudio realizado por la Comisión Europea se concluye que la mayoría de las normativas nacionales disponen de un régimen subjetivo de responsabilidad, pero se considera que pueden existir problemas para demostrar el daño, por lo que se busca aligerar la carga de la prueba. En este sentido, la propuesta de Directiva establece un régimen de exhibición de pruebas, esto es, la obligatoriedad de que los Estados miembros garanticen que los órganos jurisdiccionales tengan potestad para ordenar, a petición del demandado y previa presentación de hechos y pruebas suficientes, la exhibición de las pruebas pertinentes de que disponga. Además, se acompaña este extremo, atendiendo al propio tipo de responsabilidad (subjetiva) aplicable, de una presunción de culpabilidad en caso de incumplimiento (con ciertos matices). Asimismo, en una compleja redacción, incorpora una presunción general del nexo causal entre la culpa y el daño en determinadas situaciones. En definitiva, la propuesta de Directiva incluye una dualidad de nexos causales e impone obligaciones a los demandados en relación con la exhibición de pruebas (con importantes consecuencias jurídicas derivadas de su incumplimiento). La transposición de estas reglas, en ocasiones, no será una tarea sencilla en ciertos Estados miembros (como el español) y se tendrán que aprobar disposiciones especiales.

ORTIZ FERNANDEZ, M., *Los sistemas Inteligentes en la propuesta de Directiva de la Unión Europea sobre responsabilidad civil*, IurisCrimPol, 2024

IV. Breve introducción a la regulación actual sobre Inteligencia Artificial en la Administración de Justicia.

No es objeto de este trabajo un exhaustivo ensayo sobre toda la regulación existente a día de hoy en el entorno nacional y europeo sobre la materia que nos concierne, pues el mismo excedería con mucho los límites que se tratan de analizar. Sin embargo, si se juzga conveniente la realización de un somero repaso por aquellos estadios más relevantes, con la intención de entender la preocupación que los pasos agigantados que año a año avanzamos en la materia sugiere.

1. Entorno Europeo ³²

1.1. Reglamento sobre la Inteligencia Artificial.

Esta es la normativa más completa y reciente sobre la regulación de la IA aprobada por la Unión Europea el 13 de marzo de 2024. Tras muchos meses de debate, la Eurocámara ha aprobado el Reglamento sobre la IA, respaldado por esta con 523 votos a favor, 46 en contra y 49 abstenciones, lo que explicita la voluntad de generar una regulación omnicomprendensiva sobre una realidad tan presente.

Esta nueva ley busca establecer un marco jurídico uniforme, en particular en lo que respecta al desarrollo, la comercialización y la utilización de la Inteligencia Artificial de conformidad con los valores de la Unión Europea. Es decir, busca regular tanto la introducción en el mercado como la puesta en servicio de determinados sistemas de IA, en consonancia con la intención de la UE de erigirse como referente y líder mundial en el desarrollo de una IA segura y ética. En esta norma se ha definido explícitamente desde una perspectiva jurídica, por primera vez, el concepto de Sistema de Inteligencia Artificial como aquel sistema basado en una máquina diseñado para funcionar con distintos niveles de autonomía, 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 información de salida, como predicciones, contenidos, recomendaciones o decisiones, que puede influir en entornos físicos o virtuales³³.

Además, uno de los puntos más novedosos de esta normativa es que el Reglamento de la IA sigue un enfoque fundamentalmente basado en el riesgo. Es decir, categoriza los diferentes sistemas de IA dependiendo del nivel de riesgo que estos supongan, diferenciando entre:

³² Al efecto, la obra de DE HOYOS SANCHO, M. *El uso jurisdiccional de los sistemas de Inteligencia Artificial y la necesidad de su armonización en el contexto de la Unión Europea*, 2023; realiza un gran análisis sobre la propuesta de armonización en materia de Inteligencia Artificial en el entorno europeo, siendo consciente de la necesidad de generar una normativa europea armonizada, en especial, en relación con los sistemas denominados “de alto riesgo”.

³³ Artículo 3.1 Reglamento de Inteligencia Artificial.

1. Prácticas de IA prohibidas, por vulnerar derechos fundamentales, como manipular personas o grupos vulnerables³⁴. Su uso está estrictamente prohibido salvo en situaciones muy excepcionales.
2. Sistemas de IA de alto riesgo, permitidos en el mercado europeo con requisitos obligatorios y evaluación de la conformidad *ex ante*.³⁵ En esta categoría, se deben evaluar y reducir los riesgos previstos, mantener registros de uso, ser transparentes y precisos, y contar con supervisión humana. Además, los ciudadanos tendrán en todo caso derecho a presentar reclamaciones sobre los sistemas IA y a recibir explicaciones sobre las decisiones basadas en estos que afecten a sus derechos.
3. Sistemas de IA de riesgo limitado, en los que los requisitos son menos estrictos.
4. Sistemas de IA de riesgo bajo o mínimo, para los que no se establecen regulaciones específicas.

Además, la regulación prevé multas administrativas de diversa cuantía (de hasta 35.000.000 euros o, si el infractor es una empresa, de hasta el 7% del volumen de negocio total anual mundial del ejercicio financiero anterior, si esta cuantía fuera superior), en función de la gravedad de la infracción.

1.2. Libro Blanco sobre la Inteligencia Artificial de la Comisión Europea.

Publicado definitivamente en 2020, este Libro Blanco creado por la Comisión Europea es el primer documento de estas características que trató de generar una guía estratégica en relación a cómo abordar el nuevo paradigma que la IA plantea.³⁶

En este texto, la Unión Europea abraza el nuevo avance, tratando de fomentar que los estados traten de apostar porque la UE sea un “líder mundial de la innovación en la economía de datos y sus aplicaciones”. Para ello, presenta dos bloques diferenciados: el desarrollo de un “ecosistema de excelencia” y el desarrollo de un “ecosistema de confianza”.

El “ecosistema de excelencia” hace referencia al marco político de cada Estado, previendo que se armonicen los esfuerzos de cada Miembro en base a unos estándares europeos. Para ello, el Libro Blanco juzga necesario aumentar la inversión en el sector de la IA, pues, según los datos que presenta, en 2016 se dedicaron tan solo 3.200 millones de euros a investigación IA, mientras que Asia dedicó 6.500 millones y EEUU 12.100 millones.

³⁴ Ejemplo de estas son aquellas prácticas con potencial para manipular personas mediante técnicas subliminales que trasciendan su consciencia o que aprovechen la vulnerabilidad de grupos vulnerables concretos, como son los menores o las personas con discapacidad, para alterar de manera sustancial su comportamiento de un modo que es probable que les provoque perjuicios físicos o psicológicos a ellos o a otras personas.

³⁵ Permitidos en el mercado europeo siempre que cumplan determinados requisitos obligatorios. Cabe destacar que se regula para los proveedores de IA el deber de implementación de sistemas de gestión de riesgos asociados a estos sistemas de IA de alto riesgo.

³⁶ Al efecto, la obra de LORENZO PÉREZ, C. *Inteligencia artificial en la Administración de Justicia: Regulación española y marco europeo e internacional. Proyectos desarrollados por el Ministerio de Justicia de España*. Centro de Estudios Jurídicos, 23 de marzo de 2022; en la que se realiza un somero análisis de las necesidades españolas, así como un repaso del marco normativo a nivel internacional y nacional de la utilización de Inteligencia Artificial en la Administración de Justicia.

El “ecosistema de confianza” prevé el estudio y la mejora en el ámbito legal para adaptar los sistemas nacionales a una nueva realidad, que debe seguir ofreciendo a los ciudadanos seguridad para que adopten las nuevas aplicaciones de Inteligencia Artificial. Así como seguridad jurídica a las empresas y organismos para que innoven y apuesten por ella.³⁷

Como reflexión, cabe reconocer el esfuerzo, ya en épocas tempranas, de la UE de adelantarse a una realidad que irremediamente iba a asentarse en la sociedad. Sin embargo, este esfuerzo se ha criticado por entenderse demasiado abstracto, pero carente de soluciones. En apenas 32 páginas, la UE pretendió establecer las bases de este nuevo paradigma, pero haciendo caer todo el peso adaptativo a los Gobiernos nacionales quienes, inexpertos, en muchas ocasiones no pudieron adoptar las sugerencias de este informe de 2020.³⁸

1.3. *Reglamento General de Protección de Datos*³⁹ (en adelante, RGPD)

Aquí podemos encontrar la base regulatoria de la Inteligencia Artificial en la UE.

En concreto, es importante destacar su artículo 22, en el que se reconoce el derecho a los ciudadanos a que las decisiones tomadas sobre ellos no se basen únicamente en un tratamiento de los datos automatizados, o en la elaboración de perfiles e inclusión del individuo en uno de ellos. Este derecho, sin embargo, no es un derecho absoluto, sino que se excepciona en casos como cuando esté expresamente reconocido por el Derecho comunitario o cuando haya sido expresamente consentido por el ciudadano.

En definitiva, podemos ver como ya en 2016 el legislador ya erigió este Reglamento como un límite al uso de las tecnologías y de protección de los derechos del ciudadano.

³⁷ <https://datos.gob.es/es/noticia/el-libro-blanco-de-la-inteligencia-artificial-aboga-por-la-excelencia-y-la-confianza>

³⁸ Se considera imprescindible referenciar, al efecto, el artículo de DE HOYOS SANCHO, M., *El Libro Blanco sobre Inteligencia Artificial de la Comisión Europea: Reflexiones desde las garantías esenciales del proceso penal como “sector de riesgo”*, Revista Española de Derecho Europeo, pág. 1 a 76, 2021.

³⁹ 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

2. Entorno Nacional

2.1. Constitución Española: ¿Exigencia de un Juez con rostro humano?⁴⁰

Nuestro sistema judicial está construido en torno al concepto de legitimidad democrática. Esto significa que España, sistema democrático, está sustentado en tres pilares fundamentales provenientes de la necesaria división de poderes que Montesquieu ya adelantó en su tiempo. El Poder Judicial, uno de estos pilares básicos del Estado de Derecho, corresponde a los jueces por decisión del pueblo (pues su poder para juzgar y ejecutar lo juzgado deriva de la Constitución, Carta Magna aprobada por el pueblo mediante su voto). Además, derivado de esta potestad jurisdiccional se regula todo el procedimiento de acceso a la carrera judicial, el cual, como se explicará más tarde, trata de asegurar que las personas que acceden a esta posición reúnen los requisitos necesarios para ejercer tan importante cargo.

En ninguna norma, ni nacional ni de nuestro entorno se prevé, expresamente, que los jueces han de ser personas (lo que, por otro lado, parece perfectamente normal, al resultar otro caso impensable en el momento de la redacción de estos textos). Sin embargo, y entre otros, en nuestra Constitución encontramos una referencia clara a los Jueces y Magistrados en el artículo 117.1 CE (“La justicia [...] se administra [...] por Jueces y Magistrados”). Pese a que no existe una referencia expresa a la necesidad de que estos Jueces y Magistrados sean personas humanas, la jurisprudencia -escasa- ha interpretado que el juez al que se refiere la Constitución (al menos en Alemania, único caso de nuestro entorno en el que se ha generado el litigio⁴¹), ha de ser un ser humano, lo que parece excluir a priori a las personas jurídicas como jueces y, de la misma manera, a las máquinas inteligentes.

V. Inteligencia Artificial como sustituto del magistrado.

1. Introducción:

En los últimos tiempos, parece que la realidad va a poder superar a la ficción. Se dice que en los últimos 100 años la tecnología ha avanzado más que en los últimos 1000, lo que plantea, desde luego, incertidumbres y retos a los que tanto los ciudadanos como los legisladores tendrán que enfrentarse en las próximas décadas.

⁴⁰ Se pregunta RIVERO ORTEGA, R, en su obra *May Robots replace civil servants?*, Derecho Digital e Innovación N° 16 (abril-junio), 2023; sobre no solo la posibilidad, sino también sobre la constitucionalidad de esta nueva propuesta. Trata sobre la insuficiencia en la regulación de la automatización y los algoritmos, sobre el potencial régimen jurídico de los robots y, de forma extensa e informada, sobre los límites constitucionales al reemplazo de los empleados públicos por los algoritmos en el entorno europeo.

⁴¹ Para Alemania, v. NINK, D. (2021), *Justiz und Algorithmen: über die Schwächen menschlicher Entscheidungsfindung und die Möglichkeiten neuer Technologien in der Rechtsprechung*, Ed. Duncker & Humblot, Berlin

Sin duda, uno de estos avances es la Inteligencia Artificial, y su potencialidad para sustituir los trabajos de miles de personas. Esta tecnología, la más puntera hasta la fecha, promete generar una revolución tecnológica que dará lugar a una nueva era, pero... ¿puede afectar a la Administración de Justicia? Y en su caso ¿puede sustituir, en todas sus tareas, a un juez?

Comenzaré este epígrafe advirtiendo que no son pocos los debates que este tema sugiere, ya sea por razones o motivos éticos⁴² o por el miedo al cambio de paradigma, unido a la desconfianza que la delegación de este tipo de decisiones en las máquinas plantea.

En las páginas que siguen trataré de abordar este problema, tanto desde un planteamiento teórico (cabida constitucional) como desde un planteamiento práctico (analizando casos de uso reales y potenciales).

Resta decir, analizando la cuestión, que parece que, teóricamente y aunque con limitaciones, esta potencial sustitución podría tratarse de una opción viable. La IA actual (ChatGPT-4) ya es capaz de resolver cuestiones -si bien simples- sobre derecho, incluso, aprobar exámenes de aptitud de prestigiosas facultades⁴³. También ha sido capaz de aprobar el examen SAT, obteniendo mejores resultados la versión ChatGPT v4 (ahora mismo, disponible para todo aquel que la quiera utilizar, pero de pago) que la versión gratuita de la misma herramienta, lo que nos da una idea de los pasos agigantados a los que la IA avanza con cada nueva actualización.⁴⁴

Pese a este relativo buen desempeño, los jueces se han declarado, en su mayoría y de manera firme, en contra de que la Inteligencia Artificial sustituya a jueces y magistrados.

Uno de los defensores más tenaces de esta posición es Manuel Marchena, presidente de la Sala Segunda del Tribunal Supremo, quien ha reiterado, en diversas ocasiones, su postura al respecto. En una conferencia denominada “Inteligencia Artificial y proceso penal: ¿hacia una justicia robótica?” que tuvo lugar en el Parlamento de Canarias el día 12 de julio de 2022, el magistrado indicó que la Inteligencia Artificial no puede sustituir

⁴² Ejemplo de esto, la Sentencia dictada por la Corte de Distrito de La Haya, el 5 de febrero de 2020 sobre el Sistema de Indicación de Riesgos -SyRI- que decidía, unilateralmente, qué familias se encontraban en riesgo de exclusión social, concediendo a estas una ayuda y denegándola a aquellas familias que no cumplieran sus requisitos. Con el tiempo, se descubrió que esta IA estaba siendo discriminatoria en contra de las personas de color o las madres solteras. Tras un análisis exhaustivo, el tribunal dictaminó que este algoritmo resultaba incompatible con el derecho a la vida privada, con el derecho al proceso debido y a otros derechos fundamentales.

⁴³ En 2023, se puso a prueba la capacidad de ChatGPT para generar respuestas en los exámenes de graduación del grado en derecho tanto de la Facultad de Minnesota, como en un MBA de la Facultad de Wharton. Los docentes de ambas Universidades calificaron los exámenes a ciegas, aprobando a ChatGPT en ambos casos, aunque con una calificación mediocre (pero superior en el examen de la Escuela de Negocios). En el examen en derecho, que capacita al estudiante para ejercer como abogado en el Estado de Minnesota, ChatGPT obtuvo un C-, inferior a la media del resto de estudiantes (B+), pero en la escuela de negocios obtuvo un B, B-; equivalente a un 8/10.

En ambos casos, los docentes indican que su capacidad de desarrollo y de generar argumentaciones firmes era superior a la media, pero que tuvo “errores sorprendentes” en preguntas y matemáticas básicas.

<https://cnnespanol.cnn.com/2023/01/26/chatgpt-aprueba-examenes-leyes-derecho-trax/>

⁴⁴ ChatGPT4 obtuvo una nota dentro del percentil 10% superior de la cohorte, mientras que la versión original de ChatGPT aprobó dentro del percentil 10% inferior.

<https://es.cointelegraph.com/news/chatgpt-v4-aces-the-bar-sats-and-can-identify-exploits-in-eth-contracts>

la labor de un juez, pues la Justicia ha de tener “un rostro humano”. Indica que “nuestra jurisprudencia tiene que ser cambiante, adaptarse a cada caso concreto. ¿Cómo se podría sustituir a ese juez que va a poder valorar la mirada del testigo o la firmeza con la que defiende su relato? No basta con afirmar si es técnicamente posible sustituir al juez o no, sino plantearse si eso hace una sociedad más justa. En ese caso mi respuesta es absolutamente negativa.” Para luego aseverar, “impartir justicia es algo más que decir cuál es el derecho aplicable. Una Justicia robotizada sin selección sería una Justicia que se alejaría del valor constitucional”⁴⁵.

Además, en caso de que fuera técnicamente posible la creación de una IA tan desarrollada como para poder sustituir a los jueces en su labor... ¿tendría esta nueva posibilidad cabida en nuestro sistema jurídico? Y, en su caso, ¿en concepto de qué?

Un juez-robot sería, desde el punto de vista físico, una máquina, pero esta no es una clasificación que nuestro sistema jurídico haya regulado con anterioridad. Es más, en nuestro Código Civil no parece que este híbrido encaje de manera sencilla en ninguna de las clasificaciones actuales. No se trata de una persona física (arts. 29 y 30 CC), pues no ha sido concebido, ni ha nacido. Tampoco es una persona jurídica (art 35 CC), ni es animal, ni cosa (al menos, no sólo es cosa de acuerdo con el artículo 333 CC⁴⁶). Ante esta dificultad existen diferentes doctrinas que defienden variadas propuestas. Un sector cualificado desea categorizar a estos jueces-robot como cosas⁴⁷ (concretamente como bienes muebles, como serían un horno o una tostadora), mientras que otra pretende crear una figura nueva, que sería la de “persona electrónica”⁴⁸.

En caso de que se siguiera lo pretendido por la segunda doctrina, sería necesaria la aprobación de un “Estatuto jurídico del robot”, en el que se decidiera jurídicamente cuál es la naturaleza de este y dónde se fijarán las funciones que puede y no puede cumplir, así como la responsabilidad en que puede incurrir la máquina cuando, como consecuencia de la realización de sus funciones, cause daños a personas y cosas. Siguiendo esta lógica, en el caso de que se declarara a este tipo de máquinas “persona electrónica”, se habría de dotar a estos robots de personalidad jurídica propia⁴⁹, con todo lo que ello conlleva.

⁴⁵ No son pocos los defensores de esta postura, jueces y magistrados que consideran que, de acuerdo con su experiencia, sentido común, intuición o sabiduría pública, pueden interpretar el tono, los gestos, las expresiones faciales y el lenguaje corporal para detectar, entre otros, si un testigo miente o dice la verdad. Lo cierto, sin embargo, es que estos jueces y magistrados carecen, en numerosas ocasiones, de formación en psicología del testimonio y neurociencia y que, por tanto, dicha supuesta capacidad para hallar verdades con una simple mirada es, cuando menos, cuestionable. ¿Será acaso más certera la Inteligencia Artificial entrenada para reconocer micro expresiones que jueces y magistrados con décadas de experiencia?

⁴⁶ Como defiende Juan-Luis Gómez Colomer, en el Capítulo 5 de su Manual “El Proceso Penal ante una nueva realidad tecnológica europea”, pág 171.

⁴⁷ Lacruz Mantecón en *Robots y personas. Una aproximación jurídica a la subjetividad cibernética*, 2020; entre otros, considera que jamás los robots podrán estimarse como personas, ni humanas, ni siquiera como jurídicas, alegando para ello que en los robots falta tanto inteligencia, entendida como conciencia moral, y autonomía, entendida como libertad.

<https://idibe.org/doctrina/personas-cosas-animales-robots/>

⁴⁸ Posición defendida por el Parlamento Europeo en su resolución de 16 de febrero de 2017, y, en España, principalmente por Moisés Barrio Andrés.

⁴⁹ Cabe destacar que, de forma no vinculante, la Unión Europea ya ha recomendado seguir este camino en la Resolución del Parlamento Europeo de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho Civil sobre robótica:

Parece ser que si se les reconociera a las máquinas naturaleza de persona electrónica se estaría creando un sujeto jurídico (quizá se daría un proceso similar al que está ocurriendo en estos momentos con la regulación de los derechos de los animales), lo que significa que quizá estos tendrán derechos que los humanos debemos respetar, pero... ¿cuáles serían estos derechos? ¿y cuáles serían sus deberes? Además, al poseer personalidad jurídica propia... ¿deberían ser tomados como responsables por sus acciones?⁵⁰

En los epígrafes que siguen trataré de ahondar en las posibilidades de progreso, sin olvidar el presente, para conseguir una aproximación de lo que el futuro de la tecnología nos depara en el ámbito de la Justicia.

2. ¿Qué es un Juez y cuál es su labor?

Un juez es la autoridad máxima en un Órgano o Tribunal de Justicia.⁵¹ El artículo 117.1 de la Constitución Española establece que los jueces integran el Poder Judicial. Son definidos como independientes, inamovibles, responsables, y sometidos únicamente al imperio de la ley.

Todas estas condiciones reconocidas constitucionalmente otorgan (o pretenden otorgar) al juez la capacidad de juzgar libremente, sin influencias ni condicionantes externos. Por ello, para cumplir con su misión esencial de administrar Justicia, su labor debe estar en el acatamiento de la ley.

Además, se accede a esta posición a través de la carrera judicial, pues el hecho de ser juez supone aptitudes, capacitación y condiciones personales que aseguren el desempeño que la función exige. Para ello, entre otros, el artículo 301 de la Ley Orgánica del Poder Judicial dispone que el acceso a la categoría de juez habrá de estar basado en los principios de mérito y capacidad. Para garantizar dichos principios se establece que el acceso a esta posición se adquiera a través de un cuidadoso proceso de selección, en el que se garantice con objetividad y transparencia la igualdad de todos aquellos aspirantes que reúnan las condiciones necesarias de acceso al cargo.

“f) crear a largo plazo una personalidad jurídica específica para los robots, de forma que como mínimo los robots autónomos más complejos puedan ser considerados personas electrónicas responsables de reparar los daños que puedan causar, y posiblemente aplicar la personalidad electrónica a aquellos supuestos en los que los robots tomen decisiones autónomas inteligentes o interactúen con terceros de forma independiente”.

⁵⁰ Pese a que en la realidad todavía no existen intentos de regulación concreta en la materia, la ciencia ficción ha generado obras maestras que tratan los dilemas éticos que este futuro generaría. Entre otros, cabe resaltar el dilema ético que supone la potencial libertad y personalidad jurídica de un robot en “el Hombre Bicentenario”, de Isaac Asimov, como se ejemplifica con maestría en este fragmento:

“-El no sabe qué es la libertad. Es un robot.

-Papá, no lo conoces. Ha leído todo lo que hay en la biblioteca. No sé qué siente por dentro, pero tampoco sé qué sientes tú. Cuando le hablas, reacciona ante las diversas abstracciones tal como tú y yo. ¿Qué otra cosa cuenta? Si las reacciones de alguien son como las nuestras, ¿qué más se puede pedir?

-La ley no adoptará esa actitud -se obstinó el Señor, exasperado. Se volvió hacia Andrew y le dijo con voz ronca:- ¡Mira, oye! No puedo liberarte a no ser de una forma legal, y si esto llega a los tribunales no sólo no obtendrás la libertad, sino que la ley se enterará oficialmente de tu fortuna. Te dirán que un robot no tiene derecho a ganar dinero. ¿Vale la pena que pierdas tu dinero por esta farsa?

-La libertad no tiene precio, Señor -replicó Andrew-. Sólo la posibilidad de obtenerla ya vale ese dinero.”

⁵¹ Definición de acuerdo con la página web ConceptosJurídicos.com, en el área de Derecho Procesal.

Como vemos, la ley trata de generar un férreo sistema de garantías, blindando la calidad y conocimientos de aquellos aspirantes que desean pasar a formar parte de la columna vertebral de nuestro Estado de Derecho. Tanto los jueces como los magistrados son pilares fundamentales en nuestro sistema judicial, contribuyendo a la Justicia y el Estado de Derecho en nuestro país.

Por ello se les exige que sean independientes, inamovibles, responsables, que estén sometidos únicamente al imperio de la ley, que hayan pasado todas las pruebas de aptitudes necesarias para ser reconocidos como personas capacitadas y que juren (o prometan) guardar y hacer guardar fielmente y en todo tiempo la Constitución y el resto del ordenamiento jurídico, lealtad a la Corona, administrar recta e imparcial justicia y cumplir sus deberes judiciales frente a todos.

Y ahora, debemos cuestionarnos... ¿puede una Inteligencia Artificial cumplir dichas garantías? ¿Puede una IA jurar ser independiente, inamovible, responsable, así como guardar fielmente la Constitución y el resto del ordenamiento jurídico?⁵² ¿puede una IA ser leal? Y en caso de no serlo, ¿puede ser tomada como responsable?

Pese a que se generan muchas preguntas, existen pocas respuestas. Parece que, como la situación no ha llegado todavía, los diferentes legisladores se abstienen de contestar a dichos interrogantes. Sin embargo, en países como Estonia (en un caso de éxito que analizaré con posterioridad) o en China (cuyos datos son algo menos fiables), parece que no ha existido problema moral ni técnico para delegar, definitivamente, un sector del poder judicial a unas máquinas preparadas por el Gobierno para resolver ciertos tipos de juicio (en Estonia, juicios por cuantía inferior a 7000 euros, y siempre pudiendo ser

⁵² Parece claro que una IA no puede jurar, pero, al no poder jurar, tampoco puede faltar a su juramento (algo que sí ocurre, en ocasiones, con los humanos, incluidos los jueces). Pese a que no puede jurar unas normas externas, una IA puede ser programada, no pudiendo faltar a su programación interna. Mientras un humano puede actuar en contra de su conciencia o incluso perderla, una IA no podrá contradecir ni actuar en contra de su programación. A este respecto, el autor Isaac Asimov ha teorizado al efecto en su Obra Maestra “Yo, Robot”, en la que plantea un futuro en el que humanos y robots conviven, estando todos los robots programados para, en toda situación, actuar conforme a Tres Leyes de la Robótica:

1. Un robot no hará daño a un ser humano, ni por inacción permitirá que un ser humano sufra daño;
2. Un robot debe cumplir las órdenes dadas por los seres humanos, a excepción de aquellas que entren en conflicto con la primera ley;
3. Un robot debe proteger su propia existencia en la medida en que esta protección no entre en conflicto con la primera o con la segunda ley.

En esta obra, Asimov plantea una serie de relatos cortos en los que dichas leyes internas entran colisión unas con otras, y que podrían servir de ejemplo de cómo un robot reaccionaría, en el futuro, cuando dichas normas superlativas impresas en su Código más interno entraran en colisión con las diferentes situaciones en la realidad.

En específico, es interesante el desgarrador relato del Hombre Bicentenario (Isaac Asimov) en el que dos humanos ordenan a un robot libre desmontarse a sí mismo. El robot ha de obedecer, por orden y en todo caso, las Leyes de la Robótica. En este supuesto, entran en colisión las Leyes de la Robótica segunda y tercera, debiendo prevalecer la segunda y por ello, comenzando este robot, con pena e impotencia, a destruirse a sí mismo.

Quédate quieto —le ordenó el alto, y le dijo al otro—: Podemos desmontarlo. ¿Alguna vez has desmontado un robot? —¿Nos dejará hacerlo? —¿Cómo podría impedirlo? Andrew no tenía modo de impedirlo si le ordenaban no resistirse. La Segunda Ley, la de obediencia, tenía prioridad sobre la Tercera ley, la de supervivencia. En cualquier caso, no podía defenderse sin hacerles daño, y eso significaría violar la Primera Ley. Ante ese pensamiento, sus unidades motrices se contrajeron ligeramente y Andrew se quedó allí tiritando.

El Hombre Bicentenario, Isaac Asimov.

apelada su decisión frente a un Juez de carne y hueso) y liberar a los jueces humanos de parte de la gran carga a la que se ven sometidos. Estonia forma parte de la Unión Europea y desde que este tipo de jueces-robot entraron en vigor en 2019, nadie ha puesto en duda la constitucionalidad de la medida, lo que podría sentar un precedente positivo para su instauración en el resto de la Unión.

Sin embargo, mi opinión personal es otra. Tras reflexionar, considero que una máquina, por lo menos en su estado actual, no podría tener cabida como juez o magistrado de acuerdo con la regulación constitucional del momento. Esto se debe a que, aunque de acuerdo con la literalidad de la ley a los jueces se les exigen unas cualidades que quizá podríamos redefinir para las máquinas (independencia, inamovilidad, responsabilidad, y sometimiento únicamente al imperio de la ley), la finalidad del legislador al establecer estas máximas era, de manera clara, definir unas características intrínsecamente humanas y de necesaria protección.

Por el momento y por definición, no creo que una máquina pueda ser independiente, desde el momento en el que es programada por alguien, que decide qué datos y qué ordenes introducir en su corazón. Además, creo que la independencia es una cualidad humana abiertamente relacionada con la libertad. Somos independientes porque somos libres, y somos libres porque somos humanos. Libres para conocer el mundo que nos rodea y decidir en consonancia con nuestros valores, y libres para juzgar el bien y el mal, la justicia de la injusticia, de acuerdo con nuestras experiencias, nuestro fuero interno y nuestra humanidad.

No creo que una máquina pudiera alcanzar, nunca, esta cualidad a la que muchos humanos no consiguen llegar nunca, y que aspirantes a jueces y magistrados tardan tantos años en conocer e interiorizar.

Por otro lado, la inamovilidad también me genera dudas. ¿puede una máquina ser inamovible?

Para los jueces de carne y hueso, esta garantía significa el derecho a no ser removidos de su cargo salvo por las causas legalmente establecidas. Aporta al sistema jurídico firmeza, y permite a los juzgadores decidir de acuerdo con su criterio, sin miedo a represalias.

Pero... ¿cómo se desarrollaría esta garantía en estos supuestos juzgadores de cable y metal?

No creo que pueda existir una esencia que el ordenamiento jurídico juzgue conveniente mantener en las máquinas, ese éter preciado y frágil que sí existe con los jueces y magistrados actuales. ¿qué sería considerado vulnerar esta garantía en las máquinas? ¿su reprogramación? ¿quedan estas, entonces, blindadas, una vez utilizadas por primera vez en el sector público? ¿no genera esto un peligro extremo para nosotros, los ciudadanos? ¿y no cambian acaso los jueces de parecer con el paso de los años, con el nuevo ajuste de la sociedad? ¿cómo ajustaríamos estos hechos en la máquina, sin vulnerar la tan querida inamovilidad?

El tema de la responsabilidad se tratará con posterioridad, pero es sin duda uno de los pozos sin fondo más desoladores y profundos del momento. La propia UE duda, y no termina de decidirse sobre cómo tratar la responsabilidad de las IAs. La incertidumbre es máxima. ¿puede una máquina ser responsable? ¿puede ser castigada por decidir de acuerdo con lo que alguien programó en su interior? ¿y cuál sería el supuesto castigo?

En este punto parece evidente una distinción fundamental e importante: las máquinas, tal como las estamos desarrollando, se están convirtiendo en expertas en fingir ser como nosotros. Pero nosotros, los humanos, no podemos caer en el engaño de que realmente lo son.

Una máquina, por muy sincera, real y humana que parezca, no deja de ser un conjunto organizado de bits. No siente, no tiene ideas propias, no es capaz de entender más allá que emulando lo hecho por humanos. Una máquina no puede ser castigada, no se le puede hacer razonar sobre un sentido más profundo de la Justicia y la injusticia que aquello que esta infiere de los datos que se le han nutrido. Una máquina no tiene memoria, no tiene sentido del bien y del mal y no entiende nuestras relaciones sociales, interpersonales, ni humanas.

Una máquina, por lo menos, tal como las conocemos por el momento, es un objeto inerte, un objeto no vivo. Es más, en muchos casos, una IA como las propuestas se trata de algo intangible, que está y no está a partes iguales, un conjunto de bits, de pulsos eléctricos que danzan al son del programador. Y yo me pregunto, ¿cómo podemos hacer responsable a algo que no existe?

La última de las características se refiere al sometimiento absoluto al imperio de la ley.

De nuevo, el mismo dilema. ¿a qué está sometido un robot?

La respuesta parece clara: a aquello que sus programadores hayan decidido que lo esté.

Este es un punto realmente complejo, pues creo que, si a una máquina tan solo le proporcionáramos los textos legales, podría convertirse en un monstruo. En un primer momento ni siquiera los procesaría. Se necesita un largo y arduo trabajo para generar un robot capaz de entender el lenguaje, de procesar las palabras y, finalmente, de entender el sentido de las frases. Para ello, la forma de hacerlo es a través del acceso a internet, que ya está introduciendo en la máquina sesgos y conocimientos ajenos a la ley.

Pero incluso aunque pudiéramos, de alguna manera, mantener estos sesgos fuera de corromper a la máquina, nos encontraremos con el siguiente dilema. La máquina no sería conocedora de nuestra sociedad. No entendería cómo funciona nuestro sistema ni las relaciones interpersonales; pero tampoco conocería sentimientos como la compasión o, retóricamente, la humanidad⁵³.

⁵³ Se genera entonces la cuestión. ¿Puede una máquina ser compasiva? ¿puede una máquina tener la cualidad de la “humanidad”? ¿no se trata acaso de una paradoja irresoluble?

Entonces, estas cualidades, tan humanas, que los jueces de carne y hueso poseen al ser seres sociales que se han criado en un ambiente en el que estos valores existen, no estarían presentes en sus fallos.

¿Es posible una Justicia sin humanidad?

3. Inteligencia Artificial como evaluador y juzgador. Frío, pero imparcial.

Se ha planteado, en ocasiones, la posibilidad de instaurar un “Juez IA⁵⁴”, es decir, la posibilidad de que la IA reemplace al juez en la toma de decisiones⁵⁵. Si bien existen casos de éxito en los que aquellos procesos más sencillos han sido, en mayor o menor medida, automatizados⁵⁶, aquello que preocupa a los estudiosos sobre este problema es la capacidad de crear una IA capaz de funcionar de manera adecuada frente a los “casos difíciles” y que, además de resolverlos, sea capaz de razonar su resolución (para cumplir con la obligación de motivar las sentencias que impone el artículo 120.3 de la Constitución Española).

Lo que caracteriza a estos “casos difíciles” es precisamente que se trata de situaciones en las que es complejo encontrar una solución jurídica que no implique el sacrificio de un bien considerado moralmente valioso⁵⁷. Por ello, en estos casos, el necesario proceso para pasar desde una situación jurídica controvertida hasta su resolución pasa por múltiples dificultades y decisiones parciales, llenas de cuestiones de hecho y de Derecho vinculadas a reglas y principios que no tienen una solución directa, sino para los que habitualmente caben, por lo menos, dos interpretaciones razonables. Es decir, en estos casos podemos encontrar al menos dos fuerzas en tensión: la primera, de carácter “subsuntivo”, que implica elementos de juicio preestablecidos (es decir, las normas potencialmente aplicables al caso, los criterios jurisprudenciales, las líneas doctrinales...) que se enfrentan a una fuerza de carácter “distintivo”, que trata de extraer de las circunstancias particulares del caso motivos para exceptuar o modular la respuesta preestablecida para el caso. Esta segunda parte, estrechamente vinculada con los valores de la sociedad de la que el juez es parte, parece generar dilemas (posiblemente irresolubles) sobre cómo la Inteligencia Artificial puede aprender dichos conceptos, que el juez conoce al ser un humano integrado en la sociedad, y poder juzgar necesarias excepciones atendiendo al caso concreto⁵⁸. Por ahora, la forma en la que se están construyendo los algoritmos que

⁵⁴ GÓMEZ COLOMER, J.L., *Problemas legales del juez robot desde una perspectiva procesal y orgánica*, pág. 163 a 180. Del mismo modo, en GÓMEZ COLOMER, J.L., *Ética, robots y proceso: sobre los límites en el uso de la inteligencia artificial* y MARTINEZ ZORRILLA, D., *El juez artificial: ¿próxima parada?*.

⁵⁵ BONSIGNORE FOUQUET, D., *Sobre Inteligencia Artificial, decisiones judiciales y vacíos de argumentación*.

⁵⁶ Casos que serán desarrollados en el apartado correspondiente y entre los que podemos destacar el programa piloto de “jueces robot” en Estonia, especializados en conflictos de baja cuantía y con posibilidad de recurrir la decisión ante un tribunal humano o PROMETEA, capaz de generar propuestas de fallo en Argentina.

⁵⁷ GARCÍA OBANDO, P.A., AGUIRRE ROMÁN, J.O., PABÓN MANTILLA, A.P. *Los casos difíciles como colisión entre premisas. Un intento de explicación desde la teoría de la argumentación de Perelman*, 2009.

⁵⁸ Por ejemplo, pese a que la conducción por encima del límite permitido de velocidad puede conllevar una infracción, el tratamiento del caso podría ser entendido de manera diferente en el supuesto de un individuo que excede el límite de velocidad al haber apostado dinero en unas carreras callejeras ilegales, que en el de

podrían, en teoría, reemplazar al juez, tienen un carácter marcadamente retrospectivo que genera dudas sobre su capacidad para responder de manera adecuada a las características propias de un caso novedoso y diferente a aquellos datos que la IA posee en su sistema.

Otro punto interesante que valorar es tratar de definir si es posible que una IA pueda ser programada de forma tal que sea capaz de emular el proceso mental que jueces y magistrados realizan ante casos difíciles. Es decir, indagar sobre si una máquina puede “pensar”, en cierta forma, o por lo menos realizar un proceso lógico asimilable al proceso de pensamiento humano y, por ello, ser entrenadas para realizar dicha labor.⁵⁹

Algunas de las principales diferencias entre el pensamiento humano y el “pensamiento” que pueden realizar las IA no son solo respecto a capacidad, sino también respecto a cómo ambos organizan la información. Si bien parece que las máquinas pudieran simular un comportamiento humano, no pueden llegar a duplicarlo, pues realmente no procesan la información de la misma manera que nosotros.

Un ejemplo de cómo los pensamientos humanos y de las IA funcionan de manera diferente podría ser, por ejemplo, el reconocimiento de una caligrafía determinada.

Un humano acostumbrado a leer la letra del puño y letra de otra persona la reconoce, pues identifica patrones en las diversas letras, tendencias respecto a la separación entre las palabras, etc. Una IA será capaz de reconocer la caligrafía de una persona, pero el procedimiento será completamente diferente. Mientras el humano interpreta líneas, curvas y separaciones, la única opción para la máquina es desagregarlo todo en píxeles, es decir, disolver dichas líneas y tendencias para proceder a otorgar un valor numérico a cada píxel y, mediante un procedimiento estadístico, determinar si la posición relativa de los píxeles se corresponde con un grado de confianza suficiente como para afirmar que han sido escritos por la misma persona. Como se ve, el resultado en ambos casos puede ser muy similar (el reconocimiento de una persona como el escritor de un texto), pero el trasfondo es diferente. La IA ha perdido todo el contexto, todas las líneas, curvas, inflexiones, para solo centrarse en los píxeles. Hoy en día, la resolución de casos difíciles podría presentar los mismos problemas.

Además, la finalidad de la IA (fuerte) tal como la entendemos hoy en día es funcionalista y utilitarista, es decir, ser eficaz en la resolución de los problemas que está programada para resolver. Por ello... ¿puede la IA acompañar sus decisiones de razones significativas en los casos difíciles? Parece que lo que la IA actual está capacitada para hacer es encontrar patrones dentro de los casos que pertenecen a su haber, recopilar y agrupar “locuciones argumentales”, mediante palabras clave o similitudes y generar una vinculación en el que las circunstancias C1, C2, C3... estarían asociados a una respuesta R (por ejemplo, absolver, condenar...) y argumentando dichas circunstancias mediante ciertos argumentos A1, A2, A3... pero estos no serían sino un mecanismo de búsqueda

aquel individuo que conduce con el mismo exceso de velocidad, pues su mujer, en el asiento del copiloto, acaba de ponerse de parto con su primer bebé. En ambos casos la infracción existe, y por lo tanto habría de conllevar sanción, pero quizá el Juez valora de manera diferentes las circunstancias que rodean cada uno de los casos como diversas a la hora de realizar su fallo.

⁵⁹ Al efecto, BARONA VILAR, S. *La seductora algoritmización de la justicia. Hacia una justicia poshumanista (Justicia+) ¿utópica o distópica?*, El cronista del Estado Social y Democrático de Derecho N° 100, pág. 36 a 46.

entre los argumentos anteriormente expuestos por jueces, que la máquina estaría utilizando por analogía, sin entrar a valorar el fondo del asunto en concreto.

Sin embargo, resolver un conflicto complejo es, en muchas ocasiones, mucho más que un ejercicio rutinario de adjudicación de la razón a aquella parte que tenga de su lado la ley (si fuera tan sencillo, en muchas ocasiones no existiría un conflicto). Por ello, “hacer justicia” parece conllevar algo más que adjudicar, pues supone generar juicios sobre lo que las personas *deben* hacer. Por ello, cuando el juez ofrece razones para fundamentar su fallo, su propósito último es generar una descripción de la pauta de actuación adecuada conforme a un correcto entendimiento del Derecho y de sus valores, es decir, es en muchas ocasiones un juicio de moralidad (alguien ha actuado bien y alguien ha actuado mal), de acuerdo con los cánones de legalidad y siempre dentro de un Derecho razonablemente justo que incorpora en su seno los cánones básicos de moralidad racional del contexto social, susceptibles de ser compartidos por todos.

He aquí la cuestión fundamental. ¿Puede esta tarea ser realizada por una IA?⁶⁰

Cómo se ha argumentado, parece que no por la IA tal como la conocemos hoy en día, pues esta necesitaría entender nuestro sistema social y, de alguna manera, formar parte de él para incluirlo en sus razonamientos. Sin embargo, existe un nuevo prototipo de IA denominado IA superinteligente que podría, en el futuro cercano, conformar la nueva generación de jueces y magistrados⁶¹.

Este tipo de IA no sólo sería capaz de resolver problemas sencillos con base en datos, sino que llegaría un paso más allá. Sería capaz de entrenarse a sí misma, liberándose (teóricamente y en el mejor de los casos), de alguna manera, de los sesgos y limitaciones humanas, y poseería la capacidad de entender de una manera más profunda y similar a la humana nuestras creencias, culturas y hábitos sociales, al interactuar con los humanos de una manera más orgánica que la actual⁶². Este tipo de IA, entonces, contaría con los beneficios de la máquina (rapidez, eficiencia en el dominio de datos, capacidad de acceso a toda la información disponible en Internet en segundos) pero sin las limitaciones de esta (incapacidad de razonar como los seres humanos, necesidad de ser entrenada con datos humanos depurados, incapacidad de interpretar de forma compleja las emociones y de tener la empatía y el estándar social como un factor a tener en cuenta en sus decisiones).

Esta nueva herramienta sería el paso siguiente en el desarrollo de la IA, pero plantea, sin duda, diversos desafíos.

- Desalineación de objetivos: Si la IA Superinteligente desarrolla objetivos que no están alineados con los valores humanos, podría tomar acciones que sean perjudiciales para la

⁶⁰ Ver en los Anexos la opinión del propio ChatGPT al efecto.

⁶¹ Dilema planteado por BELLOSO MATIN, N., en *Algoritmos predictivos al servicio de la Justicia: ¿Una nueva forma de minimizar el riesgo y la incertidumbre?*, Revista da Faculdade Mineira de Direito, V.22 N.43; NIEVA FENOLL, J., *The loss of digital control: are we heading towards a judicial dystopia?*, Editorial La Ley, 2023; BARONA VILAR, S., *Datization of Justice (Algorithms, Artificial Intelligence and Justice, The beginning of a great friendship?)* Rev. Boliv. De Derecho N° 36, págs. 14 a 45, 2023.

⁶² <https://www.ibm.com/es-es/topics/artificial-superintelligence>, 2024

humanidad. Este riesgo se agrava si la IA encuentra formas de resistir o evitar intentos de ser apagada o controlada.⁶³

- Retroalimentación y pérdida de control: La posibilidad de que una IA se auto-mejore rápidamente podría llevar a un crecimiento incontrolado de sus capacidades, lo que dificultaría aún más el control humano.
- Desplazamiento laboral: La automatización avanzada podría desplazar a una gran cantidad de trabajadores, exacerbando la desigualdad económica y social. Esto podría conducir a disturbios sociales y aumentar la brecha entre ricos y pobres.⁶⁴
- Ciberseguridad: Una IA Superinteligente podría superar los sistemas de seguridad actuales, comprometiendo la privacidad, los datos sensibles y la infraestructura crítica a nivel global.
- Decisiones Éticas: La IA podría tomar decisiones que afecten profundamente la vida humana sin una comprensión adecuada de la moralidad y la ética humana.⁶⁵

Para enfrentarnos a dichos desafíos, hemos de seguir una hoja de ruta, que la Unión Europea ha marcado en su Libro Blanco para el futuro de la Unión Europea, como sigue:

- Investigación en IA Segura: Desarrollar métodos para garantizar que la IA actúe de manera segura y controlable, incluso cuando supere la inteligencia humana. Esto incluye técnicas de alineación de objetivos y robustez frente a fallos y manipulaciones.⁶⁶
- Supervisión y Regulación: Crear marcos regulatorios internacionales que supervisen y controlen el desarrollo y despliegue de IA superinteligente para prevenir usos indebidos y garantizar la seguridad.

En definitiva, la generación de una IA Superinteligente podría traer beneficios sin precedentes a la humanidad, pero también conlleva riesgos y desafíos significativos. Para abordar estos riesgos de manera efectiva, es crucial un enfoque multidisciplinario que combine la investigación técnica con consideraciones éticas, regulatorias y sociales. Parece que la colaboración internacional y la participación pública serán fundamentales para garantizar que el desarrollo de la IA Superinteligente beneficie a toda la humanidad y no ponga en peligro su futuro, aunque si seremos capaces o no, técnicamente, de desarrollar este tipo de Inteligencia Artificial, sigue siendo una incógnita.

⁶³ Distopía de la que ya encontramos ejemplos en la ficción, como en la saga “Terminator”, de James Cameron.

⁶⁴ GÓMEZ CARDOSA, D. *Ámbitos de afectación de la IA en el mercado laboral y las habilidades*, 2023. RUIZ RAMÍREZ, E. *Avances y desafíos de la inteligencia artificial en el mundo laboral. Una visión desde el sector público y privado*, 2024.

⁶⁵ https://www.elconfidencial.com/tecnologia/novaceno/2023-06-02/dron-inteligencia-artificial-mata-operador-simulacion_3658110/

Un robot militar se rebela por primera vez y mata a su operador en una simulación, tras juzgar que este era un obstáculo para cumplir su misión de ataque y destrucción de objetivos.

⁶⁶ BERRYHILL J.; HEANG K.K.; CLOGHER R.; *Hola, Mundo: La Inteligencia Artificial y su uso en el Sector Público*, Documentos de trabajo de la OCDE sobre Gobernanza pública núm. 36, 2020.

4. Análisis de caso: Jueces robot en Estonia y PROMETEA en Argentina.

A continuación, pasaré a analizar un par de casos en los que, frente a la incertidumbre, los diversos Estados han dado el paso (siempre de la mano de desarrolladores privados) para conseguir involucrar de manera pseudocompleta a la IA (fuerte) en los procesos judiciales. Sin embargo, el problema planteado *supra* sobre los casos difíciles se mantiene, pues estas IA tan sólo están programadas para la resolución de casos sencillos y directos, en el que el mayor peso del caso es matemático (mediante operaciones aritméticas) o de aplicación directa de la ley (sin que existan varias soluciones posibles, o ninguna).

4.1. Programa piloto de “jueces robot” en Estonia, especializados en conflictos de baja cuantía y con posibilidad de recurrir la decisión ante un tribunal humano.⁶⁷

Estonia, un país relativamente joven y relativamente pequeño es, contra aquello que pudiera parecer, uno de los países más avanzados tecnológicamente de nuestro entorno⁶⁸. Quizá esto se debe a las ingentes cantidades de dinero que cada año el Gobierno dedica a I+D y al desarrollo de las TIC. Especialmente, en 2018 el gobierno estonio contrató a Ott Velsberg⁶⁹ para encabezar un nuevo proyecto en varios Ministerios que simplificara los servicios que estos ofrecían a la población.

El proyecto sin duda más ambicioso y polémico de los llevados a cabo por este proyecto tuvo que ver con la implantación de “jueces robot”⁷⁰. La finalidad de los mismos era la resolución de todos los juicios menores (definiendo como tales aquellos en los que existan disputas por cuantía de 7.000 euros o inferior) que pudieran darse en el país a partir de 2020. Así, encargando estos conflictos a la IA se permitía acelerar decenas de casos atrasados a los que los jueces y secretarios judiciales no podían hacer frente en el momento, por las altas cargas de trabajo a las que estaban sometidos.

El procedimiento para resolver este tipo de casos era sencillo: ambas partes debían cargar sus documentos y toda la información relevante en una plataforma digital. Con base en esta información y la legislación de Estonia y de la UE, la IA emitía una decisión que siempre podía ser apelada ante un juez humano⁷¹. Es importante destacar que la máquina no tomaba decisiones propias; se limitaba a buscar en su sistema casos similares al

⁶⁷ <https://www.thetechnolawgist.com/2019/06/12/estonia-se-prepara-para-tener-jueces-robot-basados-en-inteligencia-artificial/>

⁶⁸ Por ejemplo, más de 1.3 millones de ciudadanos poseen en activo una tarjeta de identificación nacional que les posibilita la realización de una gran cantidad de servicios públicos en línea, como el voto electrónico.

⁶⁹ Ott Velsberg es el actual Jefe de Datos de Estonia, así como un referente mundial de la Revolución IA.

⁷⁰ GÓMEZ COLOMER, J.L., *Problemas legales del juez robot desde una perspectiva procesal y orgánica*, pág. 168 y CÁRDENAS KRENZ, R. *¿Jueces robots? Inteligencia Artificial y Derecho*, Revista Justicia & Derecho.

⁷¹ GÓMEZ COLOMER, J.L. “Problemas legales del juez robot desde una perspectiva procesal y orgánica”. Capítulo 5, *El proceso penal ante una nueva realidad tecnológica europea*, 2023 págs. 163 a 175.

propuesto, identificar el presente y aplicar la misma solución que los jueces habían decidido en situaciones anteriores.

Este sistema, que hoy en día todavía está en uso, parece efectivo y certero para ciertos casos, pero realmente no resuelve el problema principal: la automatización de la resolución de los casos complejos, pues son estos los que requieren la mayor parte del tiempo y atención de jueces y magistrados, así como son estos en los que los volúmenes de información a valorar son superiores. Por otro lado, también es dentro de los casos difíciles en los que el juez ha de interpretar dentro del margen de discrecionalidad que permite la ley, y por ello, aquel es el momento en el que los sesgos propios del humano detrás del juez pueden interferir con la necesaria impartición de Justicia, defecto humano que la IA podría solucionar.

Como reflexión sobre esta posibilidad, encontramos un punto débil: la incapacidad de la IA para discernir la verdad de la mentira, y la justicia de la injusticia. Dejar a merced de las partes los documentos que incorporar al caso podría sesgar la respuesta, al no ser estos documentos verificados directamente por un humano. En caso de que alguna de las partes introdujera documentos falseados, beneficiando sus intereses, la IA no sería capaz de identificarlo, sino que los tendría en cuenta como información verídica de cara a su resolución. Además, a día de hoy, el tono en el que los diversos escritos estén presentados también puede jugar un papel en la resolución del caso, lo que no ocurriría con un juez, quién sería capaz de entrever las intenciones de una de las partes y contrarrestar dicho efecto.

4.2. PROMETEA, análisis aplicado al caso del uso de la IA como asistente del Sector Público.⁷²

Desde 2018 se utiliza en Argentina la Inteligencia Artificial para asistir en la resolución de causas criminales. Su origen proviene de una rara colaboración entre el Ministerio Público Fiscal de la Ciudad de Buenos Aires, juristas de reconocida trayectoria y sólida formación, y técnicos y programadores especializados en Inteligencia Artificial.

El objetivo de esta herramienta pionera fue, en palabras de su fundador “terminar con la injusticia de que un mismo ciudadano, con el mismo problema, recibiera dos respuestas diferentes por parte del Estado. Es durísimo cuando esto pasa, y viola un principio básico, que es el de la igualdad”.⁷³

PROMETEA es un sistema de *software* que tiene como cometido principal la automatización digital de procesos de tareas reiterativas y la aplicación de IA para la elaboración predictiva de dictámenes jurídicos en los casos análogos para cuya solución ya existen precedentes judiciales reiterados. En Argentina, cada vez que se plantea un caso ante el TSJ, la Fiscalía tiene como deber emitir un dictamen jurídico con el objetivo de proponer una solución jurídica. Este dictamen no es vinculante para los jueces, pero sí obligatorio en cada causa.

⁷² <https://mpfciudad.gob.ar/institucional/2020-03-09-21-42-38-innovacion-e-inteligencia-artificial>

⁷³ ESTEVEZ, E.; LINARES LEJARRAGA, S., FILLOTTRANI, P., PROMETEA, Transformando la Administración de Justicia con herramientas de Inteligencia Artificial, 2020.

Además, un 66% de los casos que ingresan en el TSJ tratan sobre “amparos habitacionales”⁷⁴. Existen cuatro tipos principales de amparos habitacionales, presentando cada uno de ellos sus características y requerimientos propios. Sin embargo, una vez probado que el individuo se encuentra dentro de una de esas cuatro situaciones, la jurisprudencia ha resuelto repetidamente en el mismo sentido en la gran mayoría de las ocasiones⁷⁵.

Del mismo modo, existen otras clases de amparos reiterados cuya solución jurídica está estandarizada, entre las que se han de destacar las causas vinculadas a ejecuciones fiscales. Estos casos se tratan de situaciones en las que el gobierno solicita la ejecución judicial de una multa y el litigio no tiene sentencia firme todavía. En estos casos, el TSJ deniega automáticamente la ejecución judicial.

Así pues, del total de casos que se presentan anualmente frente al TSJ, un 70% se encuentran englobados entre amparos habitacionales y causas vinculadas a ejecuciones fiscales.

Aprovechando este alto porcentaje, se ha entrenado a la IA PROMETEA para que pueda predecir la solución de estos casos en segundos, con tan sólo introducir el número de expediente en el sistema. Lo que hace la máquina es “leer” todas las actuaciones previas del supuesto, y realizar un análisis de las palabras clave del expediente comparándolo con los patrones de los precedentes judiciales de la base de datos para, en menos de quince segundos en promedio, predecir cuál es la solución que debería ser adoptada y generar una propuesta de fallo.

Acto seguido, se genera en pantalla para el usuario el modelo de dictamen que debería utilizar, y se realizan unas breves preguntas (mediante un chatbot por texto o por voz) para adecuar el modelo propuesto al expediente concreto. Acto seguido, la Fiscalía recibe el expediente en papel, listo para presentar.

Por el momento, la solución jurídica es siempre supervisada por un funcionario antes de firmar, aunque los estudios realizados en 2018, que le aportaban un 96% de tasa de acierto, han hecho levantar voces acerca de la posibilidad de eliminación del control a posteriori.

En la actualidad esta herramienta está en pleno funcionamiento en diversas instituciones, tanto en la Ciudad de Buenos Aires como en otras provincias del país. Sin embargo, no parece haber dado el salto internacional, y aunque ha sido probado por ciertos organismos (como la Corte Interamericana de Derechos Humanos), ninguna otra organización ha

⁷⁴ Estos son los casos en los que un ciudadano argumenta no poseer vivienda alguna y reclama al gobierno de la ciudad que le provea de una, sustentándose en el derecho constitucional a la salud integral (art. 20 de la Constitución de la Ciudad Autónoma de Buenos Aires) y a una vivienda digna (art. 31 de la Constitución de la Ciudad Autónoma de Buenos Aires).

⁷⁵ Si se prueba que se trata de la clase uno, personas solas con buen estado de salud, la jurisprudencia reiteradamente deniega el amparo. Para la segunda clase de personas, que son aquellos adultos sanos con niños, los precedentes judiciales ordenan que el gobierno provea una ayuda monetaria. La tercera y cuarta clase se trata de individuos con discapacidad y personas mayores de 60 años, respectivamente. En estos casos, la jurisprudencia indica que el Gobierno debe proporcionar al peticionario una vivienda digna integral, ya sea cediéndole una vivencia o suministrándole una ayuda monetaria suficiente para pagar un alquiler.

dado el paso en tratar de adquirir dicha tecnología. Además, los datos sobre su uso y su eficacia no son públicos hoy en día.

En este punto, es obligada la reflexión. PROMETEA está, hoy en día, generando el 70% de los informes preceptivos que se le presentan al TSJ de Argentina. De estos, el 96% son correctos, y no han de ser modificados por los funcionarios. Esto significa que, *de facto*, en el 67,2% de las causas que llegan al TSJ de Argentina, el informe que supuestamente debería haber sido generado por funcionarios del Estado, capacitados para ello, ha sido creado por una IA y no modificado por un humano. ¿Nos encontramos frente a una dejación de tareas del Estado para con sus ciudadanos? ¿y como esto afecta a sus derechos? ¿quién es responsable en caso de un dictamen incorrecto que perjudique los derechos del ciudadano?

Especialmente en los casos de amparo habitacional, el juez del TSJ dicta directamente en función del informe preceptivo proveniente del fiscal. Al estar este generado por PROMETEA ¿están los derechos de los ciudadanos argentinos en manos de la IA?

Los promotores de esta disruptiva IA defienden su utilidad mediante máximas como “PROMETEA hace en dos minutos lo que a la Corte le llevaba tres días⁷⁶”. Sin embargo... ¿es el ahorro de tiempo el bien final buscado por el derecho? ¿es este un fin que debemos buscar por encima de la eficacia, la humanidad, o la protección íntegra de los derechos de cada ciudadano?

No cabe duda de las diferencias que pueden existir entre un trabajo “de dos minutos” frente a un trabajo de “tres días”. ¿Estamos cambiando rapidez por reflexión? ¿es algo que realmente queremos (y debemos) hacer?

Por otro lado, esta IA tan solo es capaz de basarse en casos anteriores, previamente registrados en su sistema. Sin embargo, ningún caso va a ser exactamente igual a uno pasado, sino que existirán matices, problemas, diversas situaciones, que generarán discrepancias entre los supuestos. Ceder la capacidad de generar informes completos y proyectos de fallo a una máquina significa aceptar que esta, incapaz de discernir entre el bien y el mal, va a tener que tomar decisiones y ponderar situaciones, basándose para ello en una amalgama de casos similares (que no iguales) dictados por múltiples personas con diversas opiniones, voluntades y, además, en diferentes momentos temporales (lo cual puede generar una gran divergencia en los fallos, pues el derecho va muy unido a la sociedad del momento).

Por otro lado, la generación de estos informes se está tomando como un acto interno de la Administración, por lo que no se está informando al ciudadano afectado de que su informe, que teóricamente debía haber realizado un funcionario capacitado, ha sido generado en apenas minutos por una máquina. Tampoco se le informa de que este mismo informe ha incluido un proyecto de fallo que, en el caso de los amparos habitacionales, va a ser aceptado por el juez de manera semiautomática. ¿Se están respetando de forma suficiente los derechos de los ciudadanos? ¿y se está cumpliendo con el necesario principio de transparencia que sobrevuela todo lo relativo a la IA?

⁷⁶ Lucía Bellocchio, representante de la CIDH en una entrevista para La Nación, 2017.

Ni en la Constitución Argentina ni en la Constitución propia de la Ciudad Autónoma de Buenos Aires existe regulación específica que exija la necesidad de informar al ciudadano de que en su sentencia ha participado, de forma directa o indirecta una IA (al contrario de lo que ocurriría en la Unión Europea, donde el destinatario de la resolución tiene, en todo caso, derecho a conocer si la IA ha participado en la resolución de su sentencia y en qué medida.⁷⁷). Además, no existe en este país recurso alguno específico para denunciar la no conformidad por este dictamen realizado por la IA, sino que la falta de conformidad habrá de alegarse contra, en su caso, la posible sentencia una vez dictada, con el aumento de coste y de tiempo que estos suponen para el ciudadano.

Del mismo modo, los dictámenes producidos por PROMETEA no son infalibles. ¿Afecta de alguna manera a la igualdad de los ciudadanos que estos en algunos casos sean generados por la máquina, y en otros por funcionarios? Y de la misma forma, respecto al juez ¿resulta discriminatorio que en ocasiones el juez admita y reproduzca el proyecto de fallo generado por la máquina, y que en otros lo desheche y lo realice por sí mismo? ¿debería en alguno de estos casos dar cuenta de lo sucedido al ciudadano? ¿tiene este derecho a conocer quién ha dictado la sentencia definitiva?

Y en caso de que el juez resuelva en contra del dictamen generado por PROMETEA... ¿debería esta situación ser comunicada al justiciable? ¿podría ser motivo de recurso que el juez falle en contra del dictamen preparado por la máquina? ¿y debería ser especialmente motivada la decisión del juez de apartarse de lo sugerido? Es decir, ¿puede el juez fallar en contra de lo indicado por la máquina, diseñada esta para ser justa, y con acceso a información mucho más amplia de aquella que dispone el juez? En caso de respuesta negativa a esta pregunta ¿acaso no se está desvirtuando precisamente la labor de los jueces, relegándolos a la opinión (supuestamente informada e imparcial) de una máquina? ¿y cómo afecta esto a la independencia judicial?

En definitiva, PROMETEA pretende facilitar las tareas de jueces y funcionarios. Para ello, trata de automatizar tareas, fundamentando sus decisiones en grandes volúmenes de información provenientes de fallos y dictámenes anteriores. Sin embargo, esta forma de proceder tan solo la hacen capaz de resolver casos similares a aquellos que se encuentran en su base de datos, y para que dos casos sean similares, estos han de ser, sin duda, sencillos (pues la incorporación de detalles generaría discrepancias entre ambos). Estos casos sencillos son aquellos que para las personas requieren, de igual manera, menos tiempo y carga, por lo que su automatización no estaría siendo tan útil como se pretende. Sin embargo, aquello que se pone en juego es mucho. Las IA tienen la tendencia de sufrir alucinaciones⁷⁸, que en el ambiente de la Justicia podrían ser fatales, generando

⁷⁷ En España, esta situación sería impensable, pues existen diversos cuerpos normativos que cubren esta posibilidad. Entre ellos, destacamos el artículo 50.1 del Reglamento de la IA.

⁷⁸ “Alucinación” es un término utilizado en el entorno de la IA para referirse a aquellas situaciones en las que esta, al realizar un juicio de probabilidad para determinar cual es la respuesta más plausible frente a una cuestión determinada, termina decidiendo que lo es una respuesta que, fácticamente, es incorrecta. Estas situaciones se dan habitualmente cuando la información disponible en Internet sobre el tema concreto es confusa, diversa, limitada o simplemente inexistente. Las IAs habitualmente están programadas para dar, siempre, una solución a lo planteado, por lo que realizarán dicho juicio de probabilidad y responderán sin avisar de ninguna forma al usuario de la posibilidad de estar generando una respuesta incorrecta. La mala práctica en el uso de la IA y el exceso de confianza han conllevado, en ocasiones, a que se cometan errores por parte de los humanos al utilizar la información propuesta como cierta, sin realizar las comprobaciones pertinentes.

soluciones que, aunque a primera vista podrían ser plausibles, fueran intrínsecamente injustas.

También nos enfrentamos aquí al problema de los sesgos. Que la IA se nutra de casos anteriores parece necesario, pero puede estar creando en la máquina un sesgo implícito que esta reproducirá en sus informes, perpetrando la injusticia.

En definitiva, muchos son los interrogantes y pocas las respuestas para esta novedosa máquina que, tras años funcionando en Argentina, parece no haber podido despegar y que sin duda tendría una difícil acogida en la Unión Europea, donde la transparencia es un principio fundamental que respetar con preponderancia sobre la eficacia.

VI. Inteligencia Artificial como asistente del magistrado.

1. Introducción

Una vez analizada la dificultad para que la IA sustituya al juez o magistrado, por lo menos en un futuro cercano, queda analizar la posibilidad (mucho más actual) de que la IA pueda asistir a este en sus funciones. Quizá por prudencia o quizá por miedo, parece más fácilmente aceptable que la tecnología complemente, pero no sustituya, el sistema legal preexistente.

Sin embargo, se generan ciertas cuestiones de difícil solución. ¿Si el juez delega una parte de su trabajo en una Inteligencia Artificial, y la actuación de esta genera daños, quién es el responsable?, ¿tiene el juez que vigilar cada paso que la Inteligencia Artificial realiza por su cuenta? ¿se cumplen, en dicho caso, los objetivos de mayor eficiencia que promovieron, en un primer lugar, la delegación en este tipo de máquinas?

Pese a que no existe todavía una respuesta clara a estas preguntas, son abundantes los casos de uso en el que la tecnología e Inteligencia Artificial han sido introducidos en diversas fases del proceso, tanto en nuestro sistema jurídico como a nivel internacional -aunque en diferente medida, siendo los EE. UU. y Asia más proclives a delegar fases del proceso que en Europa no queremos automatizar-.

En el ámbito español, es un hecho que nuestro sistema de Justicia está colapsado. De acuerdo con los datos oficiales del Consejo General del Poder Judicial, la media en meses de duración de un procedimiento judicial es, para Audiencias Provinciales, de 11.6; siendo cada vez más elevada año a año (en 2016, este tiempo era de 6.4). Para los Tribunales Superiores de Justicia, la media es de 4.9 años y en el Tribunal Supremo 24.9 meses, es decir, más de 2 años.⁷⁹

Además de realmente inconveniente, el Tribunal Constitucional ha indicado que estos retrasos en la celebración de juicios (especialmente, los poco complejos) suponen una vulneración del derecho a la tutela judicial efectiva, en su vertiente del derecho a obtener una resolución en un plazo razonable.⁸⁰

⁷⁹ Consejo General del Poder Judicial

<https://www.poderjudicial.es/cgpj/es/Temas/Transparencia/Estimacion-de-los-tiempos-medios-de-duracion-de-los-procedimientos-judiciales/>

⁸⁰ Sentencia 125/2022, de 10 de octubre de 2022. Recurso de amparo 8133-2021. ECLI:ES:TC:2022:125

Los expertos indican que pueden existir dos soluciones al efecto: o aumentar el número de jueces, magistrados, fiscales, letrados de la administración de justicia y demás funcionarios, lo cual requiere de tiempo y presupuesto; o mejorar la eficiencia de aquellos que ya están en el ejercicio de sus funciones.

Es en el ámbito de esta segunda posibilidad en el que se plantea la implementación de la Inteligencia Artificial como asistente del magistrado, haciéndose cargo de tareas automatizables⁸¹ y permitiendo a los magistrados dedicar su tiempo a otras actividades de mayor valor.

2. Asistencia en la redacción de fallos y resumen de la información. Reino Unido a la cabeza y una España tímida frente al avance.

La Inteligencia Artificial ha demostrado ser capaz de procesar grandísimas cantidades de información de manera eficiente y rápida. Por ello, muchos expertos parecen acoger la posibilidad de utilizar dicha tecnología como parte del proceso judicial, al poder los robots analizar los cientos de hojas registradas como diligencias para resumir, de manera precisa, la información que el juez necesita para resolver sobre el caso. Los argumentos a favor residen principalmente en el ahorro de tiempo y recursos (aunque habría de valorarse cuánto peso tiene el ahorro de dinero al estar hablando de la impartición de Justicia) y una posible mayor imparcialidad al entender los datos, sin dejarse afectar por la intención del relato (lo que, de nuevo, puede ser un arma de doble filo, pues la resolución de un caso complejo a menudo se encuentra en los matices). Se podría, entonces, minimizar el riesgo de errores humanos debido a despistes, falta de atención, cansancio o imposibilidad material de retener todos los datos relevantes, así como podrían eliminarse los sesgos de los juzgadores (aunque quizá sustituirse por los sesgos de la máquina los cuales son, en gran parte de las ocasiones, tan solo una reverberación de los sesgos de la sociedad).

Aunque esta es una medida que ya podría ser implementada, pues existen herramientas capaces de analizar grandes textos y resumirlos de acuerdo a unos parámetros establecidos⁸², parece que, pese a que han existido intentos (como Sumar.IA, novedoso programa destinado a resumir sumarios y realizar búsquedas concretas dentro de estos el cual, tras ser testado por más de 5000 abogados el año pasado y generado una supuesta exactitud del 98,7%, nunca salió a la luz) no se ha terminado de implementar en nuestro sistema.

⁸¹ Se abre aquí, sin duda, un debate sobre cuáles son las tareas susceptibles de automatización y cuáles son las tareas de mayor y menor valor que realizan los jueces y magistrados. Todo el sistema de acceso a la carrera judicial está repleto de formalidades que tratan de salvaguardar las garantías preceptivas, así como buscar la excelencia en las personas responsables de mantener el orden jurídico. En ocasiones, las decisiones de jueces y magistrados suponen un cambio radical en la vida de las personas. Entonces, ¿qué parte de dicho proceso va a poder automatizarse? ¿Dónde se traza la línea entre qué cargas son accesorias y cuáles son fundamentales? ¿y quién va a poder decidir dónde se encuentra dicha línea?

⁸² Similar a lo que podría hacer para el público en general el Bot ChatGPT, pero en una red privada y en la que la información solo sea accesible a usuarios identificados, cosa que ya se hace en los despachos de abogados, como expondré en el correspondiente apartado.

¿Por qué no hemos adaptado una tecnología preexistente, que parece prometer una mejora directa a un problema tan complejo como es el manejo de ingentes cantidades de información? ¿No es patente que las máquinas están mejor diseñadas para resolver este tipo de problemas?

Como parece indicar la línea doctrinal mayoritaria, la reticencia proviene de que todavía no existe una regulación explícita que resuelva el problema de la responsabilidad. Como veremos *a posteriori*, en casos en los que un juez o magistrado utilice ChatGPT o similares en sus sentencias y se genere una responsabilidad derivada de un fallo incorrecto, la solución no está clara. Tampoco lo está si, a causa de utilizarlo, la sentencia es incorrecta, injusta o simplemente carece de sentido (pues cabe recordar que este tipo de lenguajes están diseñados para tener siempre una respuesta a las preguntas de los usuarios. Cuando no la tienen, se la inventan, en un fenómeno denominado “alucinaciones”⁸³, y no indican al usuario que la información no proviene de otra fuente preexistente).

Cabe destacar, como caso cercano pero diverso a lo que ha entendido razonable la Unión Europea, que, desde 2023, Reino Unido ha permitido oficialmente a sus jueces la utilización de la Inteligencia Artificial⁸⁴ (en concreto, de ChatGPT). Esta decisión resultó realmente controvertida, pero fue justificada al entender que “la utilización de la Inteligencia Artificial en toda la sociedad está en aumento, al igual que sucede en los tribunales [...] Esta guía es el primer paso dentro de un conjunto de acciones destinadas a apoyar al poder judicial en sus interacciones con la IA”. Así pues, establece una guía con cautelas y normas de “uso correcto” en la utilización de la Inteligencia Artificial, remarcando que “los jueces deben verificar la precisión de las respuestas de la IA antes de tomar decisiones que alteren el curso de la vida de las personas”. En definitiva, con esta cautela se está manteniendo la responsabilidad por fallo de la Inteligencia Artificial en los jueces, dejando a su buen juicio si utilizar la información prevista por ella o no.

Este enfoque también puede resultar controvertido. Desde luego, parece que la utilización de ChatGPT no es una solución que cumpla con las garantías que se pretenden en el ámbito de la Justicia (así como podría fácilmente generarse un problema con la confidencialidad de los datos, si jueces y magistrados introducen características de los casos concretos en el chatbot, lo que, a su vez, será un requerimiento para que la máquina pueda decidir de manera informada). Además, sin la formación tecnológica suficiente, es fácil que estos jueces y magistrados caigan en la falacia de creer que la información aportada por la máquina es acertada (aunque esto es, precisamente, lo que trata de evitar esta guía de medidas).

Por otro lado, considero que, para que la máquina pueda resultar de utilidad, esta tiene que ser puesta en contexto del caso de la manera más minuciosa posible, pues la omisión de detalles puede generar una respuesta diversa a la justa por parte de la máquina, y esta puede influenciar la decisión del juez, terminando este dictando una sentencia injusta que no habría dictado sin la recomendación de la máquina.

⁸³ <https://www.bbc.com/mundo/noticias-65606089>

⁸⁴ Artificial Intelligence (AI) Guidance for Judicial Office Holders, 12 December 2023. <https://www.judiciary.uk/wp-content/uploads/2023/12/AI-Judicial-Guidance.pdf>

Por otro lado, si los jueces no pueden fiarse de aquello que el chat les diga... ¿cuál es su utilidad real? ¿y no sería acaso más útil (si Inglaterra ha decidido que quiere introducir una IA en su sistema judicial) desarrollar una propia, cerrada y programada al efecto, en lugar del genérico ChatGPT?

En España, lejos de haber adoptado una postura similar, tan solo se utiliza la Inteligencia Artificial como medida para anonimizar la información de los documentos públicos⁸⁵, como medida para ahorrar tiempo a los funcionarios. También se ha desarrollado un buscador de información que permite, una vez subidos los documentos legales, utilizar un lenguaje sencillo tipo pregunta-respuesta para encontrar información clave en las largas páginas de sentencias. Pese a que esta tecnología existe, parece observarse una gran reticencia por parte del sector jurídico en adaptarse a su uso, lo que podría provenir, como indicaba *supra*, de una falta de regulación sobre la responsabilidad en el uso de estas nuevas tecnologías, así como una falta de formación a jueces y magistrados sobre cómo adaptarse a las nuevas posibilidades⁸⁶.

3. Análisis de jurisprudencia en el entorno español: Posición mayoritaria.

Como indicó Vicente Magro, magistrado de la Sala de lo Penal del Tribunal Supremo, en una entrevista para el diario digital Confilegal “no puede haber una jurisdicción tecnológica, como ya he escuchado en alguna ocasión. La jurisdicción tiene que seguir siendo humana. La Inteligencia Artificial tiene que asistirnos a los jueces en nuestro trabajo, no reemplazarnos”. Sin embargo, sí es partidario de que “la IA puede hacer la vida más fácil a los jueces y magistrados”.

Así, apunta que, de acuerdo con su criterio, la Inteligencia Artificial debería ser utilizada como un análisis vivo, constante sobre la Jurisprudencia dictada por los diferentes tribunales de nuestro Estado, y utilizada como un formulador de “reglas” o “principios” en los que los jueces puedan basarse para tomar sus propias decisiones, pero indicando en todo momento de qué sentencia específica han obtenido dicha información.

De esta forma, la IA podría proponer, tras ser consultada por el juez o magistrado, listados como “los 17 criterios de la valoración de la declaración de la víctima”, las “20 Reglas o principios para entender concurrente la prueba indiciaria como suficiente para dictar una Sentencia condenatoria” o “las notas características del maltrato habitual”, a ser leídos por el juez o magistrado y tenidos en consideración (o no) en función de su criterio⁸⁷. No son pocos los detractores de esta posición, que muestran preocupaciones respecto al uso de esta nueva tecnología en el ámbito legal. El mayor argumento en contra es la pérdida de contexto. Si la Inteligencia Artificial tan solo genera listas de datos que el juez

⁸⁵ <https://mjusticia.gob.es/es/institucional/gabinete-comunicacion/noticias-ministerio/ia>

⁸⁶ Imaginemos qué ocurriría si un juez utiliza la herramienta para localizar palabras clave en un texto, utilizando para ello una palabra diversa a aquella con la que fue referenciada el concepto dentro del texto. Al no encontrarla en el buscador, asume que no existe, obviando dicha información y decidiendo de manera sesgada. Si bien es cierto que esta funcionalidad de la IA puede ahorrar tiempo, el Derecho no se basa en respuestas simples a cuestiones simples, sino que en muchas ocasiones trata de una historia compleja donde los matices pueden resultar clave, matices que se pierden si tan solo se acude a párrafos de un texto complejo, olvidando el contexto.

⁸⁷ <https://confilegal.com/20240504-vicente-magro-inteligencia-artificial-una-ayuda/>

aplica, está perdiendo el sentido del caso concreto, pudiendo desembocar en decisiones contradictorias o analogías incorrectas. Además, en el derecho cada caso es diverso, y precisamente la capacidad de empatizar y comprender las circunstancias sociales y personales del acusado es una habilidad del juez que, en nuestro sistema jurídico, parece querer protegerse.

4. Análisis de caso: COMPAS en Estados Unidos y Viogén en España.

4.1. COMPAS en el Caso Loomis, análisis aplicado al caso de la desigualdad y de los sesgos algorítmicos en el uso de la IA.

El Caso Loomis⁸⁸ es relevante en el estudio de la IA aplicada a la administración de Justicia por el dilema ético que presenta.

Se trata de un caso que tuvo lugar en los EE. UU. en 2013 y que puso en jaque la fiabilidad de la Inteligencia Artificial como decisor en casos judiciales y que, incluso hoy en día, ha dejado enseñanzas que se tienen en cuenta en el desarrollo de los más novedosos proyectos.

El litigio proviene de la utilización de COMPAS (cuyo acrónimo en español podría definirse como “Administración de Perfiles Criminales para Sanciones Alternativas del Sistema de Prisiones de EEUU”⁸⁹). Esta herramienta era (y sigue siendo) utilizada para predecir el riesgo de reincidencia de un individuo (que o bien ya ha sido encarcelado, y se revisa cómo proceder al final de su condena; o sobre un individuo que acaba de cometer un crimen, juzgando si resulta conveniente la privación de libertad o el establecimiento de medidas alternativas) sobre la base de un análisis complejo de información obtenida a través de una encuesta de 137 preguntas, dividida en varias secciones diferentes, así como de la información correspondiente a los antecedentes penales del individuo analizado⁹⁰.

Esta encuesta incluye preguntas como “¿alguien de tu familia ha sido arrestado alguna vez?” “y tú, ¿cuántas veces has sido arrestado?” “¿son tus amigos parte de alguna pandilla?”. A estas preguntas objetivas, se les suma otra batería de preguntas que podrían considerarse como de “pensamiento criminal”, entre las que se formulan cuestiones como “¿está de acuerdo o en desacuerdo con esta afirmación: está bien para una persona hambrienta robar comida?”. Cada respuesta se valora con una puntuación del 0-10, en función del riesgo de la respuesta. Es interesante destacar que también tiene en cuenta el lugar de procedencia del reo, el barrio en el que este habitaba antes de entrar en prisión, su raza y edad.

Por último, tras obtener respuesta a todas estas cuestiones se determina un valor promedio del riesgo, y se decide si el condenado puede salir en libertad bajo fianza, si debe ser enviado a prisión o recibir una pena alternativa (como, por ejemplo, un programa de tratamiento para la adicción a sustancias), o, cuando la persona ya está encarcelada, el

⁸⁸ State v. Loomis, 881 N.W.2d 749 (Wis. 2016), accesible online en: <https://law.justia.com/cases/wisconsin/supreme-court/2016/2015ap000157-cr.html>

⁸⁹ Y que en inglés significa *Correctional Offender Management Profiling for Alternative Sanctions*.

⁹⁰ LARSON, J.M; MATTU, S; How we analyzed the COMPAS Recidivism Algorithm, 2016. <https://fiatlexica.medium.com/compas-an-ai-tool-sending-or-keeping-people-in-jail-d9228df3a2c6>
<https://afraenkel.github.io/fairness-book/content/04-compas.html>

algoritmo es capaz de determinar si merece el beneficio de la libertad condicional o si debe, por el bienestar de la sociedad, cumplir su sentencia en la totalidad.

Esta herramienta de “*risk assesment*” ha sido duramente criticada principalmente por dos cuestiones: por ser considerada racista; y porque parece que no resulta tan fiable como en principio se pensaba respecto a sus predicciones⁹¹.

Con respecto al potencial sesgo racista, una investigación realizada en mayo de 2016 analizó tanto los resultados como las estadísticas de los casos en los que se había utilizado esta herramienta predictiva, llegando a la conclusión de que solo el 20% de aquellas personas que el programa había indicado que volverían a cometer un crimen violento realmente lo hizo. Pero, además, se demostró que el programa exageraba el riesgo de reincidencia en personas afroamericanas, teniendo en cuenta este origen como factor de riesgo superior (mientras que indicar que el sujeto era de raza blanca no sumaba puntos de riesgo). Es decir, se descubrió que, si se comparaban los resultados de una persona negra y una blanca con el mismo historial, la misma edad, el mismo género, el mismo pasado judicial y el mismo “futuro criminal”, el acusado negro tendría un 45% más de posibilidades de ser calificado como persona de riesgo por COMPAS que el acusado blanco.

Parece claro que, debido a este sesgo, el algoritmo podría estar vulnerando el principio de igualdad y de no discriminación⁹² (lo cual parecía deberse a los datos, sesgados, que la herramienta había utilizado durante su entrenamiento).

En este punto, y como paradigma de un ejemplo en el que un algoritmo puede estar mostrando prejuicios raciales, es donde integramos el Caso Loomis. En el caso *Wisconsin v. Loomis*⁹³, el Tribunal Supremo de Wisconsin dictó una sentencia en la que trataba, por primera vez hasta la fecha, sobre la constitucionalidad del uso de algoritmos informáticos que utilizan Inteligencia Artificial como fundamento del fallo de una sentencia de un Tribunal.

Los hechos del Caso Loomis son, de manera resumida, los siguientes: en 2013, Eric Loomis (varón blanco de mediana edad) fue detenido por agentes de policía del Estado de Wisconsin mientras conducía un vehículo robado implicado en un tiroteo. Loomis se declaró culpable de ser el conductor del coche, pero aseguró que no había disparado ni participado de ninguna otra manera en el crimen, todo ello con la esperanza de no tener que ingresar en prisión. Durante la vista de su juicio, en el que se debía decidir acerca de la libertad condicional del acusado, el fiscal aportó un informe elaborado por COMPAS, según el cual Loomis tenía un riesgo elevado de reincidir en actividades delictivas y de cometer actos violentos en el futuro. El informe indicaba que el acusado presentaba un “alto riesgo para la comunidad”, aunque el fiscal no aportó pruebas de por qué o en qué medida existía dicho riesgo.

⁹¹<https://mallika-chawla.medium.com/compas-case-study-investigating-algorithmic-fairness-of-predictive-policing-339fe6e5dd72>

⁹² <https://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?article=1028&context=compas>
<https://criminology.fsu.edu/sites/g/files/upcbnu3076/files/2021-03/Validation-of-the-COMPAS-Risk-Assessment-Classification-Instrument.pdf>

⁹³ Sentencia 13 de julio de 2016, *State v. Loomis*, 881, N.W.2d 749, 7532 (Wis, 2016).

En su fallo, el tribunal de instancia indicó: “usted ha sido identificado, a través de la evaluación de COMPAS, como un individuo de alto riesgo para la comunidad. Mediante la ponderación de diversos factores, hemos descartado la libertad condicional debido a la gravedad del crimen cometido y porque su historia personal, su historial de libertad bajo control y las herramientas de evaluación de riesgos que se han utilizado, sugieren que Ud. representa un riesgo extremadamente elevado de volver a delinquir”. En base a este razonamiento, el Juez impuso a Loomis una pena de 6 años de prisión y otros 5 años en régimen de libertad vigilada, basándose, tanto en el fallo como en la argumentación, en el informe de la herramienta COMPAS.

Loomis no aceptó dicha sentencia y decidió apelarla, alegando un uso inadecuado del sistema COMPAS y una vulneración de su derecho al debido proceso (*the right to a due process*). La argumentación de su defensa giró en torno a tres líneas principales:

- 1.- La defensa alegó que no había podido conocer la manera exacta⁹⁴ por la que se calculó la puntuación obtenida en COMPAS, pues dicha herramienta era propiedad privada (se determinó por el tribunal que era necesario guardar secreto sobre la misma, para proteger los derechos de propiedad intelectual de la empresa proveedora).
- 2.- Se alegó también que el uso de este software para la creación de sentencias vulneraba el derecho a una sentencia individualizada, ya que el algoritmo únicamente generaba datos basados en estadísticas grupales, que no tenían por qué ser ciertas en el caso concreto.
- 3.- Por último, la defensa alegó que el uso de esta herramienta constituía una discriminación por razón de género, pues este dato era utilizado en las evaluaciones (indicando un mayor riesgo en el caso del sexo masculino).

Con estos argumentos la defensa de Loomis recurrió al Tribunal Supremo del Estado de Wisconsin, quien, al publicar su fallo, desestimó todas estas cuestiones. A partir de entonces, encontramos el primer precedente relevante sobre la aceptación del uso de COMPAS o de cualquier otro instrumento automatizado de evaluación de riesgos por medio de la IA en la determinación de una sentencia.

En definitiva, en este fallo, el tribunal se pronunciaba sobre la **constitucionalidad** del uso de algoritmos procesados informáticamente como fundamento del fallo del tribunal de una instancia inferior.

⁹⁴ Es este, sin duda, uno de los inconvenientes más recurrentes sobre el uso de la Inteligencia Artificial, y se conoce como el fenómeno de la caja negra. Este ejemplifica cómo, una vez que una herramienta generativa como COMPAS ha sido programada, se convierte en una “caja negra”, en la que es realmente complicado (si no imposible) entender qué procedimientos se concatenan, siendo únicamente conocido el resultado final. Esto se debe a lo complejo de su configuración (ponderaciones de datos, ajustes, juicios de probabilidad...), pero sobre todo a que este tipo de recursos aprenden tanto de sí mismos al operar, como de los nuevos datos introducidos. Es decir, además de aquellas directrices que han sido voluntariamente incluidas como parte de su configuración por los programadores, las IA como COMPAS son capaces de analizar y extraer otras normas de comportamiento de la gran amalgama de datos que se les proveen (en este caso, sobre el historial de los acusados y sobre el fallo del juez en cada caso concreto), de los que estas herramientas aprenden, suman y tratan de simular. Esta opacidad y falta de trazabilidad de las conclusiones de la IA entra sin duda en colisión con el deber constitucional de motivar el fallo (que forma parte del derecho a la tutela judicial efectiva en nuestro país), lo cual parece plantear uno de los mayores obstáculos para la implantación de un potencial juez-robot autónomo en el futuro.

En el fallo final del Tribunal Supremo, se aceptaba la utilización de este tipo de herramientas argumentando que el derecho del acusado al debido proceso no fue vulnerado simplemente porque este no pudiera acceder a una explicación completa sobre el tratamiento informático de los algoritmos, pues, según indicó, la prueba de exactitud de estas herramientas y la capacidad del juez para entender su funcionamiento⁹⁵ (aunque el acusado no fuera capaz) resultaban suficientes para entender asegurado el derecho del acusado (aunque ni la exactitud de las herramientas ni la capacidad del juez para entender su funcionamiento se probó en ningún caso).

Resulta paradójico que, pese al reconocimiento de numerosas limitaciones, debilidades e inconvenientes en el sistema COMPAS y en el empleo de algoritmos como sustituto del raciocinio personal, el Alto Tribunal termine aceptando el mismo bajo la premisa de la libre apreciación de la prueba por parte del juez (lo que no parece que ocurriera en primera instancia, donde el juez tomó dicha prueba como cierta y definitiva, y con base casi enteramente en ella fundamentó su fallo).

Parece entreverse aquí una (peligrosa) tendencia a que la tecnología se tome como decisiva, como real y libre de errores, convirtiéndose en la base de las decisiones humanas, en lugar de un indicio o prueba más a valorar, siempre bajo la atenta mirada y a expensas de la decisión del juez y del derecho aplicable.

Por último, cabe destacar que no existen todavía en el derecho continental casos semejantes, pero no cabe duda de que acabarán llegando y se habrá de tomar la decisión sobre cómo interpretar las sugerencias de la IA, y sobre qué peso darles como fundamento en una sentencia. Si a estos informes generados automáticamente se les da un peso ingente (como en el Caso Loomis) de poco servirán las explicaciones o argumentos del abogado defensor, sobre el que el poder y supuesta omnipotencia de la IA pesará como una losa de la que es realmente difícil desprenderse.

Con el paso de los años, el avance tecnológico ha permitido estudiar en profundidad el funcionamiento de la herramienta COMPAS y, además del potencial sesgo racial mencionado, también se han realizado ensayos para comprobar la fiabilidad de las predicciones de esta IA. Como ocurrió en el caso estudiado, al tratarse de una tecnología, el humano medio (y de la misma forma el juez) en ocasiones asume que sus resultados son certeros, pues no estamos acostumbrados a dudar de lo que la tecnología nos dice.

Existe una investigación de ProPublica⁹⁶ de 2015 que demostraba lo prejuicioso que el algoritmo de COMPAS podía ser. Fue, a raíz de esta investigación inicial, cuando se comenzó una etapa de pruebas para tratar de estimar su fiabilidad real. En palabras de las

⁹⁵ En este punto surge la pregunta, ¿realmente el juez, en el 2013, era capaz de entender como funcionaban los comicios de la Inteligencia Artificial en su versión predictiva? ¿y ha de ser este un requisito a la hora de permitir o no su utilización? ¿hemos de comenzar a exigir a los jueces y magistrados un conocimiento profundo de la compleja tecnología utilizada para la generación automática de informes judiciales, para que estos puedan ser utilizados en el juicio?

⁹⁶ ProPublica es una agencia de noticias independiente y sin ánimo de lucro radicada en Manhattan, Nueva York. Está especializada en periodismo de investigación y es conocida por haber estado detrás de la revelación de ciertos datos y estudios controvertidos sobre grandes empresas, que no eran públicos hasta su publicación por esta revista.

investigadoras responsables del caso -Dressel y Harid-⁹⁷, antes de comenzar el experimento parecía que “por debajo de toda la conversación sobre algoritmos, se asumía que las predicciones algorítmicas eran por naturaleza superiores a las predicciones humanas”. Estas investigadoras realizaron un experimento en el que los individuos participantes en el estudio debían predecir la posibilidad de ser reincidente de un sujeto utilizando como datos para fundamentar su decisión tan solo siete características⁹⁸. Para ello, recogieron una colección de puntuaciones de COMPAS para 10.000 acusados, así como sus registros de arrestos en los dos años posteriores a la clasificación. Seleccionó al azar a 1000 de dichos acusados y tabuló los mencionados siete datos sobre cada uno (incluyendo edad, sexo y número de detenciones previas). Reclutó a 400 personas e informó a cada uno de ellos (que nunca habían recibido formación previa sobre psicología ni estadística, y los cuales eran de diferentes edades y niveles educativos) de que debían pronosticar cuáles de los 50 perfiles proporcionados volverían a ser arrestados dentro de dos años (qué es exactamente la misma información que COMPAS promete predecir). Los resultados del experimento fueron sorprendentes. Los humanos acertaron casi con la misma frecuencia que el algoritmo (entre el 63% y el 67% de las veces, en comparación con un 65% de media del algoritmo).

Tras la confusión inicial, el experimento se repitió con nuevas muestras, hasta que se llegó a la conclusión de que los elementos esenciales para poder realizar una predicción con garantías eran la edad y la cantidad de condenas previas. Así pues, y de forma muy simplificada: “si eres joven con muchas condenas previas eres clasificado como persona de alto riesgo, y si eres viejo y tienes menos antecedentes, eres de bajo riesgo”.

Este experimento fue sin duda un shock. El algoritmo utilizaba hasta 127 criterios y datos diferentes, para estimar un resultado que podría obtenerse de manera muy similar por una persona habitual, sin preparación psicológica, y tan solo conociendo un número restringido de dichos datos.

En este punto es necesario realizar una reflexión final. Por lo menos hasta el día de hoy, IA va unido, irremediablemente, a falta de transparencia. No somos conocedores de lo que exactamente ocurre dentro de una máquina, y este hecho crea, en ocasiones, indefensión, pero este principio fundamental de nuestro ordenamiento jurídico no puede ser superado por la voluntad de actuar más rápido. Además, debemos ser conocedores de que uno de los principales riesgos a los que tendremos que enfrentarnos como sociedad frente al desarrollo de la IA (especialmente en sectores tan delicados como la Justicia) es la generación de una falacia generalizada de fiabilidad de la máquina. Aunque es un hecho que las IA pueden ser una gran herramienta de automatización de tareas repetitivas, no son idóneas para la realización de todo tipo de tareas, pues hay algunas (como quizá, la medición del riesgo de reincidencia de un individuo) que no son (por lo menos, por ahora) automatizables.

Como seres humanos, una de las características que nos identifica es que somos seres sociales. Ser seres sociales significa mucho más que simplemente vivir en grupos, sino

⁹⁷ Dressel y Harid, dos profesores de informática de Dartmouth College, *The accuracy, fairness, and limits of predicting recidivism*, 2018.

⁹⁸ <https://www.science.org/doi/10.1126/sciadv.aao5580>

que conlleva una compleja red de tradiciones, modales y reglas no escritas que son imposibles (o por lo menos, extremadamente difíciles) de transmitir a una máquina. Como de manera muy visual se indica en el caso, los seres humanos estamos, en ocasiones, mejor preparados para resolver algunas actividades directamente que para preparar a las máquinas para resolverlas (somos capaces de, con tan solo siete datos, predecir el riesgo de reincidencia de una manera tan certera como una máquina con 127). Parece que esto se debe a nuestra capacidad intrínseca (y puramente social) de detectar patrones en la conducta del otro, capacidad de muy difícil traslación a una máquina, que, pese a que perfecta para realizar tareas repetitivas, no es capaz de captar matices que los humanos detectamos sin vacilación.

Este caso resulta realmente útil para reflexionar acerca de la utilidad de la IA. ¿Qué esté a nuestro alcance significa que es la mejor solución en todo caso?

Además, de aquellas personas que COMPAS había anticipado que serían reincidentes en la comisión de un crimen violento, tan solo un 20% lo fue. ¿Qué ocurre con aquellas personas que se vieron perjudicadas en sus condenas por la previsión errónea? ¿han de ser indemnizados aquellos que, entre otros, no fueron considerados para la libertad condicional por una tecnología que con el tiempo se ha considerado no fiable?

¿Y cómo se traduce esto en el futuro de la IA? Desde luego, parece que un supuesto como el caso Loomis sería impensable hoy en día en nuestro país, entre otros, por la falta de fundamentación de la sentencia, que en su momento tan sólo se basó en el informe de la IA. Sin embargo, se debe reflexionar. ¿apoyarse en un informe alegado por el fiscal no es precisamente justificar el fallo? así se considera en caso de que provenga de un fiscal. ¿Ha de ser diferente el tratamiento cuando estos provengan de un robot? ¿precisamente no deberían ser más fiables estos al provenir de una máquina especializada para generarlos? ¿resulta suficiente justificación para el fallo que la máquina, de manera informada, haya realizado dicho informe?

De nuevo, este caso genera muchas dudas, y habrá de ser tenido en cuenta a la hora de estudiar la credibilidad que se aporta a las IAs del futuro. Sin embargo, resulta muy complejo medir, en el momento, la fiabilidad de la máquina. Como ejemplo, en el caso Loomis, tan solo se conoció que COMPAS no era tan fiable como parecía en un primer momento *a posteriori*, cuando, al cabo de los años, sus predicciones no se cumplieron. ¿Pero cómo sabremos o no si se cumplen sus predicciones, si por miedo a que estas sean incorrectas, no las permitimos?

4.2. VIOGEN, análisis aplicado al caso del uso de la IA como predictor de reincidencia en casos de Violencia de Género.

VioGén es uno de los pocos casos en los que, hoy en día, se está utilizando la Inteligencia Artificial en España⁹⁹ para asistir a los magistrados en sus decisiones.

VioGén¹⁰⁰ es un Sistema de Seguimiento Integral para casos de violencia de género; una herramienta que coordina diferentes instituciones públicas competentes en esta materia, con el objetivo final de mejorar la atención y protección de las víctimas.

Todo comienza con una denuncia de la víctima, de algún familiar, de cualquier persona del entorno de la víctima o incluso con un atestado policial. En este punto, se pide a la víctima de violencia de género un relato libre, generando confianza con la víctima (siempre mujer) para que informe de todo lo que ha pasado.

El siguiente paso es realizar a la víctima de violencia de género una serie de preguntas sobre el episodio de agresión, su situación personal, si tiene hijos, el perfil del agresor y los posibles agravantes por vulnerabilidad (como, por ejemplo, la dependencia económica). Con este análisis se detectan dos cuestiones fundamentales: el riesgo de reincidencia en la agresión y el riesgo de homicidio.

El programa, basándose en casos anteriores, utiliza todos los datos aportados y genera un informe relativo a dichos extremos (riesgo de reincidencia y riesgo de homicidio). De todas formas, la decisión definitiva queda siempre en manos de los profesionales, que en caso de considerar un nivel de riesgo superior al que ofrece el sistema pueden subirlo, pero nunca bajarlo.

⁹⁹ Viogén se trata del más avanzado, pero no del único. En España han existido otros avances, como la Calculadora 988 o la textualización de vistas.

- Calculadora 988: Programa piloto en la Fiscalía del Tribunal Supremo en 2018, ha sido ya implantada en la totalidad del territorio español. Cuando un reo es objeto de múltiples condenas y cuando de conformidad con el Código Penal proceda la acumulación de las vistas, con tan solo introducir los datos la herramienta analiza todas las combinaciones posibles de acuerdo con la normativa vigente y genera lo que considera la “mejor solución”, que habrá de ser analizada por un juez antes de ser efectiva.
- Textualización de vistas: Se trata de una herramienta que permite la generación automática de transcripción de audio tomando como origen las grabaciones que han tenido lugar durante vistas y declaraciones judiciales. Al ser una Inteligencia Artificial, presenta componentes de aprendizaje de uso y de *machine learning*. Es capaz de reconocer diferentes interlocutores, acentos, diversas lenguas cooficiales y permite realizar consultas sobre lo escrito. Se ha registrado un 93% de acierto, y de acuerdo con los informes oficiales del Ministerio de la Presidencia, Justicia y Relaciones con las Cortes del Gobierno de España, esta herramienta ya ha sido implantada en todas las Comunidades del Territorio Ministerio (Baleares, Castilla-La Mancha, Castilla y León, Ceuta, Extremadura, Melilla y Región de Murcia), así como en Canarias, Andalucía, Galicia, Asturias y Madrid.

¹⁰⁰ <https://www.interior.gob.es/opencms/eu/servicios-al-ciudadano/violencia-contra-la-mujer/sistema-viogen/>
<https://www.poderjudicial.es/cgpj/es/Temas/Estadistica-Judicial/Estadistica-por-temas/Datos-penales--civiles-y-laborales/Violencia-domestica-y-Violencia-de-genero/Sistema-de-Seguimiento-Integral-en-los-casos-de-Violencia-de-Genero--Sistema-VioGen-/>

Una vez realizado dicho análisis, se cataloga en uno de los cinco niveles de riesgo propuestos: extremo, alto, medio, bajo y no apreciado. En función del nivel, se da comienzo a unas medidas de protección u otras (por ejemplo, en caso de haberse identificado un riesgo extremo, la mujer cuenta con vigilancia policial las 24 horas del día). Además, en todos los niveles de riesgo se ponen a disposición de la víctima servicios asistenciales e informativos, sobre a qué servicios acudir o sobre la posible ayuda que puede solicitar.

Este sistema, que comenzó siendo efectivo en 2007 ha alcanzado más de seis millones de valoraciones de riesgo, aunque en la actualidad tan solo tiene 77.000 casos activos, estando 31 mujeres dentro de este programa en riesgo extremo, 1.097 en riesgo alto, 11.364 en riesgo medio y 33.548 en riesgo bajo¹⁰¹.

Pese a que este sistema está resultando de utilidad, pues los informes son revisados por el juez a la hora de establecer las medidas a imponer al caso concreto, existe un fallo de concepto que está limitando su fiabilidad: el proceso tan solo se inicia tras denuncia, y, con datos a 2023, el 73,2% de las mujeres asesinadas durante dicho año nunca habían denunciado a su agresor.

Por ello, este sistema, que precisamente fue creado para identificar y proteger a las víctimas de mayor riesgo, falla en ocasiones a la hora de localizar a estas víctimas, que no entran siquiera a formar parte de su sistema de datos.

Como en el caso Loomis, debemos preguntarnos si la IA es capaz de realizar la abstracción necesaria que se requiere para estas clasificaciones, así como cabría cuestionarse cuántos datos de importancia existen en una cuestión tan compleja como una relación entre dos personas en la que potencialmente existe violencia de género, y cuántos de estos datos quedan fuera del ámbito de actuación automatizado de la máquina.

Parece que en VioGén se trata de soslayar esta deficiencia mediante un control humano que puede en todo caso aumentar el riesgo, pero nunca disminuirlo. Sin embargo, no podemos olvidar de que este tipo de situaciones son complejas para la mujer que las vive. En muchas ocasiones existe vergüenza, confusión, miedo frente a las potenciales represalias. La máquina no es capaz de medir ninguna de estas cuestiones.

La IA no es capaz (todavía) de tener en cuenta lo mucho que tiembla la voz de la víctima al tratar ciertos temas o como esta se estremece cuando otros son mencionados. No es capaz de entender el dolor en sus ojos, su mirada perdida, o cuándo el instinto por proteger a sus hijos hace que enmascare lo que realmente ocurrió.

De nuevo, quizá deberíamos reflexionar acerca de qué significa realmente disponer de la IA como herramienta, así como sobre cuáles son las situaciones en las que la automatización genera un mayor beneficio, y qué casos son tan humanos, que deben permanecer en las manos de estos.

¹⁰¹Datos obtenidos de la página oficial a 04/07/2024.

<https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/interior/Paginas/2023/190523-sistema-viogen-victimas-violencia-genero.aspx>

VII. Inteligencia Artificial como asistente del abogado.

Pese a las reticencias expuestas en el ámbito de la resolución de casos por parte de los jueces, parece que los abogados se han visto más propensos a utilizar, de una manera y otra, la Inteligencia Artificial en sus despachos.

1. Mejora de eficiencia: asistente jurídico.

Parece un hecho que los bufetes que consigan sacar partido de las nuevas tecnologías de IA podrían ofrecer sus servicios en un tiempo inferior y por un precio más económico.

Si pensamos en las tareas que más tiempo consumen en la preparación de un litigio, estas son: extracción de la estructura del caso, información relevante y significado de un enorme conjunto de documentos generados durante la fase de proposición de la prueba. Dichas tareas son precisamente aquellas que de manera más sencilla y eficiente pueden realizar las diversas Inteligencias Artificiales, aunque con la misma cautela que los jueces: resumir supone, en todo caso, eliminar gran parte del contexto y seleccionar aquellos elementos que se consideran más relevantes. Delegar completamente esta tarea en una IA supondría que el abogado comienza con unos datos reducidos (aunque este sea el objetivo), siendo conocedor tan solo de una parte de la información que una máquina, a su juicio y de acuerdo con sus datos, ha considerado más relevante.

Una vez que existe este resumen (que podrá ser realizado en segundos, ahorrando el trabajo de semanas para el abogado) será el turno de presentar la demanda, para lo cual la IA podría ser utilizada para redactar con celeridad versiones iniciales, citando jurisprudencia pertinente, o presentando argumentos y generando respuestas para rebatir o anticipar los argumentos en contrario. En este caso conviene explicar una división entre la Inteligencia Artificial abierta al uso ciudadano (cuyo paradigma es, sin duda, ChatGPT) y aquellas Inteligencias Artificiales Generativas desarrolladas por empresas, de pago y de acceso limitado a los usuarios identificados.

Este segundo tipo es aquel que los despachos de abogados están implementando, por necesidades de confidencialidad -protección de datos de sus clientes- y con el objetivo de no desvelar información a sus competidores¹⁰². Estas redes cerradas permiten, utilizando una estructura similar a la de ChatGPT, generar un sistema que, con base en unos datos

¹⁰² Recordemos que las IAs generativas a las que estamos acostumbrados utilizan la información aportada por unos usuarios como respuesta a las consultas de otros usuarios que estén relacionadas con el tema cuestionado. Esto ha llevado a grandes problemas de seguridad, pues al pedir, por ejemplo, a ChatGPT que este resumiera texto, la información aportada por el usuario para ser resumida pasaba a formar parte del “conocimiento” de ChatGPT, es decir, de los datos que podía utilizar para contestar a un nuevo usuario que realizara una pregunta relacionada con algún nombre o empresa nombrado en el texto aportado.

Uno de los fallos de seguridad más relevantes debido a esta comunicación de información entre usuarios fue que, en determinadas horas de un lunes de 2023, fue posible para algunos usuarios ver el nombre y apellidos, dirección de correo electrónico, datos de pago y los cuatro últimos dígitos del número de la tarjeta, así como la fecha de caducidad de la misma, de los usuarios que se habían registrado como socios de pago de la plataforma durante aquel día, si un usuario preguntaba específicamente por este dato. Ahora, este problema está completamente resuelto, pero resulta significativo de cómo ChatGPT y otras inteligencias artificiales generativas se nutren de todos los datos que puedan encontrar, sin someterlos al filtro social de sensibilidad, y los distribuyen a otros usuarios, creando situaciones peligrosas e ilegales. Hipertextual, Sección Seguridad, 24 de marzo de 2023

controlados aportados por aquel que crea el sistema, responda de acuerdo con dichos datos (que pueden ser normas de estilo, colores, formas de resumir, palabras clave a introducir...) generando unas respuestas que tan solo son visibles para el usuario autorizado, y que sirven para entrenar y pulir la forma en la que la IA generativa actúa.¹⁰³

Un ejemplo de la más pura actualidad de la forma en la que estas redes cerradas funcionan sería *Cocounsel*, “*the first reliable AI legal assistant, powered by GPT-4*”¹⁰⁴. Esta nueva herramienta desarrollada por Casetext permite al abogado hacer el mismo tipo de preguntas que podría hacerle un asociado, como, por ejemplo, pedirle que investigue qué han fallado los tribunales de una determinada jurisdicción en casos que presenten similitudes a aquel en que se está trabajando, a lo que la herramienta contestará generando un informe expresando todos los casos de interés, con referencias a la jurisprudencia correspondiente y explicando por qué esta es similar al caso actual. Otros ejemplos de uso de esta herramienta son la revisión de documentos, los cuales se pueden subir a la plataforma para que la herramienta los revise, y conteste a preguntas concretas sobre los mismos (manejando grandes cantidades de información en segundos) y reseñando los fragmentos de los documentos de los que ha obtenido la información, para que esta sea fácilmente verificable por el abogado¹⁰⁵.

Por otro lado, también es capaz de, introduciendo simplemente los datos y los documentos del caso, analizar cuáles serían los argumentos que proponer en el juicio, e incluso redactar un borrador de cada uno con los temas fundamentales a desarrollar. Desde luego, también es capaz de resumir, interpretar y condensar la información crítica de documentos complejos (como contratos o normativas), con una eficacia superior al 98%. Por último, en su nueva versión es capaz de analizar las cláusulas de un contrato, interpretarlas y generar un reporte de riesgos de estas, así como sugerencias de mejora.

Esta herramienta, que requiere de preparación respecto a su uso, lleva un año en el mercado anglosajón (pues es el único idioma y la única normativa que, por ahora, entiende la herramienta). Tan solo 4 meses después de su lanzamiento, esta herramienta fue comprada por Thomson Reuters por 650 millones de dólares, y está actualmente expandiéndose a Canadá y a Australia.

¹⁰³ Esta IA generativa se puede entrenar mediante la entrega de textos y del resumen correspondiente, aprobado por las normas de calidad de la firma, para que la máquina analice qué puntos resultan de relevancia y qué características ha de incluir en sus respuestas. Además, la máquina necesita recibir *feedback* de sus propias interpretaciones (lo cual se denomina “entreno”), para pulir aquellos detalles que terminan creando un mecanismo cuasi-perfecto.

¹⁰⁴ Cocounsel, el primer asistente fiable de Inteligencia Artificial para el sector legal, alimentado por ChatGPT-4.

¹⁰⁵ De nuevo cabe hacer aquí una llamada al problema de la responsabilidad. ¿Es responsable el abogado que se fía de lo que la IA provista por su despacho le indica? ¿o es responsable el programador, por los fallos que esta cometa? ¿o es quizá responsable el despacho en general? Además, ¿qué ocurriría si en juicio, la parte contraria (o el juez) descubren que parte de las sentencias citadas para sustentar los argumentos en contrario son falsas, y realmente no se corresponden con un caso que haya tenido lugar en la realidad? ¿existiría aquí algún tipo de responsabilidad especial? ¿se estaría tratando de “engañar” al juez, aunque fuera de forma inintencionada? ¿acaso deberemos regular un tipo penal nuevo para cubrir dichos casos, en los que se ha generado daño, sin voluntad, pero actuando sin la diligencia debida? ¿y dónde se encuentra aquí la línea que marca la diligencia debida?

En este momento, más de 45 grandes firmas de los Estados Unidos la están utilizando, lo que significa que tienen acceso a esta más de 50.000 abogados, aunque todavía se encuentra en una fase piloto que no es perfectamente fiable.

En las primeras encuestas realizadas, estas firmas indican que han reducido la necesidad de trabajo paralegal (o incluso juniors) en un 60%, lo que les ha hecho más eficientes con su tiempo y ha aumentado su capacidad de resolver casos con el mismo número de empleados¹⁰⁶.

Parece que todavía serán necesarios años (quizá décadas) hasta que esta herramienta sea de utilidad en nuestro derecho. Los expertos insisten en que el sistema de la *common law* (basado en el precedente judicial) permite que sea más sencillo el desarrollo de esta herramienta, y que en los sistemas continentales (basados en la interpretación de la norma) la herramienta presenta mayores problemas. Además, los mercados europeos son más pequeños, y se necesita de una millonaria inversión para “entrenar” a esta Inteligencia Artificial Generativa a interpretar el sistema jurídico de los diferentes países.

Sin embargo, y aunque parece que esta herramienta omnicomprendensiva no puede ser de aplicación inminente, existen esfuerzos en nuestra península para desarrollar una herramienta de IA que asista a los despachos.

En primer lugar, hemos de hablar del despacho de abogados Cuatrecasas, pionero en la incorporación de la Inteligencia Artificial en nuestro país.

Este gran bufete de abogados está enfocado en el desarrollo de dos tipos de trabajo: la automatización de tareas de escaso valor; y la resolución de pequeños problemas legales.

Ya en 2015 esta compañía adquirió la tecnología de SEMANTI-K, especializada en el procesamiento de lenguaje natural. Más adelante, en 2017, desarrolló una tecnología especializada en la revisión de documentos y contratos, puntero para la época. Además, este bufete está abogando, hoy en día, por incorporar un nuevo perfil profesional, los denominados *legal engineers*, que serán abogados con formación tecnológica¹⁰⁷. Para significar su compromiso con el desarrollo tecnológico, Cuatrecasas es promotor de ciertas iniciativas, como Cuatrecasas Acelera¹⁰⁸, que busca identificar proyectos maduros sobre *legal tech*, que permitan desarrollar un proyecto piloto en el despacho que automatice partes de su proceso.

En cuanto a avances concretos que el despacho haya implementado a día de hoy, podemos destacar dos.

¹⁰⁶ <https://casetext.com/cocounsel/>
<https://www.fastcompany.com/91033227/casetext-thomson-reuters-most-innovative-companies-2024>
Lawnext, As Thomson Reuters Expands Casetext CoCounsel, the AI Legal Assistant, to Canada and Australia, It Provides Details on U.S. Growth.

¹⁰⁷ Información obtenida de la entrevista a Idoya Fernández y Marta Hernández, Consejeras de Cuatrecasas en el Área de Conocimiento e Innovación, realizada por el diario legal “Líder Legal” el 20 de abril de 2023 y en Legal Today nº 974, de 27 de mayo de 2021.

¹⁰⁸ <https://acelera.cuatrecasas.com/>

Cuatrecasas Automated Knowledge Experience (CAKE)¹⁰⁹: aunque poco se conoce de esta herramienta, pues el despacho no desea hacer público su funcionamiento para preservar la ventaja competitiva que genera, se conoce que esta herramienta es fundamentalmente útil para la redacción de documentos estandarizados.

Para su desarrollo fue necesario un equipo especialista, que recopiló documentos modelo (demandas, contestaciones, escritos, peticiones...), de alta calidad, completos y ajustados a los estándares de la firma, para incluirlos en la base de datos de CAKE.

Allí, un equipo de expertos pulió aquellos detalles necesarios, anonimizando cada caso y seleccionando aquellas partes que debían ser adecuadas al caso concreto (como nombres, direcciones, *petitum*...).

Así pues, el abogado de Cuatrecasas que desea hacer uso de dicha herramienta habrá de identificarse, seleccionar la rama del derecho a la que desea acceder (penal, laboral, procesal, administrativo social o mercantil) y seleccionar, a continuación, el tipo de litigio sobre el que desea generar el documento (por ejemplo, dentro del ámbito procesal, la contestación a la demanda; o dentro del ámbito mercantil, la redacción de un acta de la junta de socios de una sociedad). Una vez seleccionado el documento, habrá de contestar a unas sencillas preguntas que se le irán planteando (como nombre y apellidos del demandado y demandante, cuantía en caso de que se trate de una reclamación dineraria o datos y número de administradores en caso de que se trate del acta de una junta de socios) y que servirán para personalizar el documento autocompletando dicha información en el borrador final. Tras ello, se podrá seleccionar los argumentos generales que CAKE sugiere como más habituales para el escrito que se pretende obtener (por ejemplo, la incorporación de un párrafo sobre la legitimación del demandante; un párrafo sobre el levantamiento de velo en contrario o una petición de suspensión de medidas cautelares, según convenga). Así, el abogado puede “personalizar” según la necesidad del caso, para obtener un escrito lo más completo posible, que más tarde solo habrá de terminar de adaptar a su caso. En apenas unos clics, CAKE es capaz de generar tanto un escrito sencillo como una demanda compleja, adaptada (a salvo de modificaciones finales) a las necesidades del caso, citando jurisprudencia, doctrina y señalando argumentos de peso con una redacción impecable.

Cuatrecasas Experto Legal IA (CELIA)¹¹⁰: se trata de una herramienta pionera que, aplicando Inteligencia Artificial generativa a los servicios legales, permite redactar escritos, hacer análisis documentales y localizar información jurídica relativa a las distintas áreas de práctica del Despacho.¹¹¹

Se trata de una herramienta desarrollada con Harvey, que utiliza como base el uso de *Large Language Model*¹¹² y está apalancado en tecnologías OpenAI y GPT4.

CELIA posee un entrenamiento jurídico específico, que se centra en el reconocimiento y asociación de palabras, lo que le permite generar textos y, sobre todo, plantear cambios que mejoran el resto de las tecnologías con las que Cuatrecasas trabaja (como CAKE). Aunque se trata todavía de un proyecto piloto, un centenar de abogados lo han probado, obteniendo sugerencias muy rápidas tanto a consultas de carácter jurídico como mejoras en contratos de más de 400 páginas generados por CAKE.

“La IA Generativa va a proporcionar un salto cualitativo a Cuatrecasas y a todo el sector. Sin duda es una herramienta de ayuda para el abogado, de copiloto, que nos permitirá incrementar nuestra aportación de valor al cliente y sofisticar aún más nuestros servicios”, apunta Francesc Muñoz, CIO¹¹³ de Cuatrecasas.

En este punto, parece necesario realizar un apunte: ¿van a seguir siendo necesarios los cientos de graduados en derecho que salen año a año, como si de una fábrica se tratara, de las diversas Universidades de nuestro territorio? ¿está la figura del *junior* condenada a la extinción, al poder ser reemplazada por la IA? ¿y cómo vamos a generar abogados *senior*, si desaparece la figura del *junior*? ¿podrá en algún momento la IA sustituir esta posición también?

Además, si comenzamos a automatizar la Justicia... ¿qué sentido tiene la figura del abogado, si los informes de las máquinas hablarán por sí solos? ¿cómo puede un humano rebatir la opinión de una máquina cuyo único propósito es generar una opinión certera sobre una situación concreta?

¿Acaso debemos replantearnos todo lo que pensábamos conocer sobre la figura del abogado?

¹¹⁰ <https://www.cuatrecasas.com/es/spain/art/cuatrecasas-sella-una-alianza-estrategica-con-harvey-para-implantar-la-ia-generativa>

¹¹¹ De acuerdo con su página web (visitada a 3 de julio de 2024), el objetivo de Cuatrecasas es seguir posicionándose entre los despachos líderes en adoptar estas nuevas tecnologías como herramienta clave para ser más productivos, aportar valor añadido a la actividad jurídica, reducir riesgos y, sobre todo, dar el mejor servicio posible al cliente.

¹¹² *Large Language Model*, denominación utilizada para la forma de funcionamiento de las Inteligencias Artificiales Generativas como ChatGPT, es decir, en formato red neuronal de cientos de parámetros que se entrenan con grandes cantidades de datos y aprenden mediante la interacción con humanos o con otras inteligencias artificiales.

¹¹³ *Chief Information Officer*, es decir, responsable de los sistemas y tecnologías de información de la firma.

2. Otros asistentes jurídicos existentes hoy en día, destinados a facilitar las tareas del abogado.

Además de esta, que podría ser la más avanzada, existen una multitud de otras Inteligencias Artificiales que los abogados están utilizando, con diferentes intenciones, y con diferentes resultados. Realizo a continuación un listado de las mismas:

Mike, el primer asesor legal virtual especializado en derecho inmobiliario español, capaz de proporcionar respuestas instantáneas a través de WhatsApp.

ALMA, asistente virtual de la Mutualidad de la Abogacía. Está muy orientada al cliente, y permite buscar cualquier producto, consultar sobre coberturas de seguros o pedir información sobre la rentabilidad.

Fiabilito, proyecto de I+D legal diseñado por un grupo de ingenieros informáticos de la Universidad Internacional de la Rioja, permite a las empresas del sector financiero realizar una primera evaluación preliminar de los productos y servicios ofrecidos por ellas que implementen algoritmos de IA, con el fin de verificar si tales algoritmos son realmente fiables.

BRYTER, plataforma no-code (sin necesidad de programar) que permite digitalizar todo tipo de procesos legales -flujos de trabajo, creación de documentos, análisis de riesgos, canales de comunicación, chatbots...-. Está especialmente indicado para despachos de abogados, asesorías jurídicas de empresas, departamentos de innovación, compliance etc.

ISSA (Inteligencia Artificial + Seguridad Social), es el asistente virtual para facilitar consultas y trámites de la Seguridad Social, y está especialmente dedicado para aquellas empresas y particulares que tienen que realizar consultas y trámites de la Seguridad Social, y que al utilizarla pueden prescindir de un abogado que les asesore.

Do Not Pay, chatbot legal que empezó ayudando a particulares de Reino Unido y Estados Unidos con sus multas de estacionamiento, y que ha ampliado su abanico de reclamaciones a aerolíneas, tarjetas de crédito... resulta útil para aquellos particulares que han de realizar reclamaciones de escasa cuantía (que no requieren, necesariamente, abogado ni procurador).

Aretha, chatbot con Inteligencia Artificial diseñado por la firma legal Ontier y la editorial jurídica Lefebvre, ofrece asesoramiento legal gratuito de Covid-19 para Pymes y ONG.

LawDroid, compañía canadiense de Inteligencia Artificial legal que quiere ayudar a abogados a automatizar tareas y conversaciones a través de tres soluciones: Recepción, Paralegal y Legal Health Check.

VIII. Conclusiones

Una vez realizado el análisis anterior, podemos inferir las siguientes conclusiones:

La presencia de la Inteligencia Artificial en nuestra vida diaria es incuestionable y su influencia futura será aún más significativa. Sin embargo, en el ámbito de la Justicia, la disponibilidad de esta tecnología no implica su idoneidad universal ni su aplicación en todos los casos.

En primer lugar, los intentos iniciales alrededor del globo (como Estonia o Argentina) han hecho patente la complejidad de integrar esta innovación en el panorama español. Como se indicó en el caso de Argentina, muchos son los interrogantes todavía por descubrir (en específico, sobre los derechos de los ciudadanos tanto a la transparencia como a un proceso debido). Pareciera ser que esta posibilidad de implementar de manera tan intensa una IA en el proceso judicial tan solo ha sido posible en dicho entorno, en el que todavía queda tanto por regular y que no debe rendir cuentas a un aparato legislativo supranacional como al que España se ve sometida con la Unión Europea. Quizá el caso de Estonia nos quede más cerca, al ser este país estado miembro y estar sometido a una regulación al efecto similar a la española. Podría parecer entonces que la falta de reclamación por inconstitucionalidad de esta especie de “jueces-robot” sería una señal positiva en aras de la implantación de este tipo de tecnología en la Justicia, pero no debemos olvidar que las labores que puede realizar esta IA son muy limitadas, en causas enteramente monetarias y, además, de escasa cuantía. Parece que la labor de la IA responsable de “juzgar” en estos casos sería más la de un autómata que se dedica a generar un *output* (cifra final) a partir de un *input* (reclamación monetaria), más que la de un juez que escucha, entiende, delibera y, finalmente, juzga.

Cuanto más casos similares afloran en las diversas partes del mundo, mayor es el interés de la UE por refinar una regulación que, hasta el momento, se torna insuficiente. Y no es de extrañar, pues no es tarea fácil la regulación de la IA en el ámbito de la Justicia.

Algunos interrogantes sobre los que considero que la UE debería reflexionar antes de generar un *corpus* legislativo completo y realmente aplicable (y no, como hasta ahora, guías o líneas de actuación a aplicar por cada Estado Miembro según su mejor saber y entender) son las siguientes:

- Transparencia. ¿Cómo y hasta qué punto?

La transparencia es uno de los principios fundamentales de la Unión, y se ha repetido hasta la saciedad en todos los documentos planteados por la UE respecto a la digitalización, tecnología e IA¹¹⁴. Pero... ¿qué significa, a nivel práctico, la necesidad de transparencia? ¿se refiere a una transparencia completa, que obligue a las empresas desarrolladoras a hacer públicos sus procedimientos? ¿y cómo impactaría esto en el

¹¹⁴ Me remito al apartado IV. 3, principio de transparencia en el entorno digital, de este mismo trabajo.

desarrollo de la IA? ¿seguiría habiendo incentivos suficientes para que estas empresas desarrollen los avances como hasta ahora?

De forma contraria, encontramos el otro extremo de la balanza: que el administrado tan solo sepa que una IA ha intervenido en la decisión tomada, pero no cómo o de qué manera. ¿Cumpliría esto la necesaria transparencia? ¿vulneraría esta actuación el derecho al proceso debido? En caso de que no se considerara suficiente, ¿qué punto intermedio es el adecuado para que, sin vulnerar los derechos de los ciudadanos, exista suficiente incentivo para la empresa privada por seguir realizando inversiones millonarias en su desarrollo?

En mi opinión, parece que el punto medio se encuentra en permitir que el administrado sepa qué partes de la sentencia han sido generadas por IA (o en qué medida se ha utilizado IA para desarrollarlas), así como hacer público en todo caso qué herramientas de IA se han utilizado. Sin embargo, esta posibilidad abre nuevas cuestiones. ¿qué recurso habrá de interponer el interesado si está disconforme con una parte de la sentencia dictada por la IA? ¿o si no está de acuerdo con los informes generados por la máquina? ¿deberá interponer los recursos habituales, con el incremento en coste y en tiempo que esto requiere, si la máquina es aquella que ha cometido una imprecisión?

Por otro lado, en el caso de que la UE así lo decidiera, el principio de transparencia podría derivar en la necesidad de conocer cómo se ha obtenido el resultado, es decir, podría exigir la trazabilidad (completa o parcial) de aquello que la máquina provee como resultado final.

La trazabilidad es un tema sensible en el ámbito de la IA. Cómo se ha comentado con anterioridad, la complejidad de las redes neuronales necesarias para que una IA completa funcionen excede, con mucho, los conocimientos técnicos e informáticos de la persona promedio. En caso de que esta explicabilidad fuera un requisito para admitir la IA en el sector de la Justicia. ... ¿se impondría a los jueces la carga de aprender acerca del funcionamiento de la misma, para juzgar si cumple con las garantías necesarias para ser utilizada en juicios?

O, en caso contrario, ¿habrán de ser los mecanismos como COMPAS o Viogén analizados por una Comisión de expertos, que determinará si son “suficientemente explicables” y si cumplen con el requisito de la transparencia?

En caso de que así fuera ¿quién sentaría los estándares para juzgar esta cuestión tan compleja? Y una vez aprobados ¿se darían por buenos en todo caso, o sólo cumplirían el requisito de la transparencia, necesario pero no suficiente para ser utilizado en juicio?

Creo que la UE deberá reflexionar acerca de qué pretende exigir. Hasta el momento, no hemos encontrado unas directrices concretas de actuación, y la falta de certidumbre está generando retrasos en el potencial desarrollo nacional. Esta cuestión, que no es menor, marcará, sin duda, el futuro del desarrollo de la IA, generando un gran impulso al desarrollo (si las perspectivas son favorables, y la IA admite y regula estos avances como queridos, posibles y deseables) o podrá retirar el interés privado (si las exigencias de transparencia son demasiado estrictas).

Encontrar el equilibrio no va a ser fácil. Creo que en este punto debemos, como sociedad, decidir si estamos dispuestos a involucrar la IA en nuestras vidas, con todo lo que ello conlleva, o si no estamos dispuestos a realizar las concesiones que este avance supone.

- Datos personales y datos sensibles

Toda IA que vaya a resultar útil en el desarrollo de una tarea se nutre de datos. Esos datos son, en su mayoría, casos anteriores a los que se le da acceso, y que la máquina analiza, evalúa y de los que aprende.

En el ámbito de la Justicia, aquellos datos a los que se le está dando acceso a la máquina son, en gran parte, datos sensibles, referidos no solo a conflictos de la persona, sino también nombres, identificación, estado civil, direcciones, números de teléfono...

Día tras día, la máquina está recibiendo esta información y almacenándola en sus servidores. Ya sea por casualidad o por malicia, existe la posibilidad de que este ingente número de datos personales lleguen a las manos equivocadas, y los derechos a la intimidad de miles de personas se vean vulnerados sin remedio. La UE debería regular sobre cómo han de ser guardados dichos datos, con qué garantías, qué personas tienen acceso y cómo reaccionar en caso de que exista una brecha de seguridad.

Además, hoy en día, los datos son utilizados en ocasiones por las empresas como moneda de cambio, y muchos de los más novedosos sistemas de IA son desarrollados por empresas privadas. ¿Están seguros con ellos nuestros datos sensibles? ¿qué medidas se van a establecer para asegurarlo? ¿vamos a poder, si así lo decidimos, vender dichos datos sensibles para entrenar a la IA? ¿y cuánto cuesta nuestra intimidad? ¿una vez dada, podemos recuperarla? ¿existe el derecho al olvido digital?.

- Responsabilidad y rendición de cuentas

Uno de los temas que considero resultan fundamentales en el desarrollo de la IA va a ser la gestión de la responsabilidad. Parece que hasta el momento no se ha encontrado una solución satisfactoria al efecto, pues, al no existir regulación vinculante, ninguno de los agentes afectados considera que la responsabilidad ha de recaer sobre él. El problema se multiplica debido a la complejidad de la trazabilidad en las IA. Una vez que un caso entra en la caja negra que resulta su interior, resulta casi imposible encontrar una línea argumental clara sobre aquello que ha ocurrido y, por ello, identificar el momento exacto en el que ocurrió el fallo (por lo que se vuelve tarea imposible cargar la responsabilidad en aquel que era responsable de aquella parte del “proceso” que ha resultado defectuosa).

Existen varias posibilidades; que la responsabilidad recaiga en el proveedor de la IA, que la responsabilidad recaiga en el usuario de la IA (por ejemplo, por mal uso), que la responsabilidad recaiga en la propia IA o una mezcla de las anteriores.

Por el momento, los proveedores de IA han tratado de evadir dicha responsabilidad mediante anuncios o *disclaimers*. ¿Es esto suficiente para considerarlos exonerados del daño?

En mi opinión, la aproximación que debería utilizar la UE a la hora de regular esta materia habrá de ser similar a la utilizada para asignar la responsabilidad fuera del ámbito de la IA, pero con matices.

Quizá sería interesante comenzar investigando dónde se encuentra el nexo causal entre acción (o falta de acción) y resultado dañoso (vulneración de derechos de un individuo debido a un fallo o actuación incorrecta en la IA).

Un claro ejemplo lo encontramos en los EE.UU., los cuales utilizaban una IA como herramienta básica utilizada por el Departamento de Asuntos de Veteranos de EE.UU.¹¹⁵, para decidir qué militares tenían derecho, una vez retirados, a recibir una ayuda del Estado. En este caso y tras una investigación, se descubrió que esta IA estaba discriminando, por defecto, a las mujeres y a los afroamericanos, otorgándoles menos ayudas que al resto de solicitantes¹¹⁶.

En este caso, el resultado dañoso era el perjuicio causado a aquellas personas discriminadas por la máquina, y la acción (o en este caso, falta de acción) resultaba la no comprobación suficiente de la falta de sesgo por los programadores de la máquina.

En este caso, podemos ver la introducción de otro concepto, vital en relación con la tecnología, que es la diligencia debida. En estas IAs complejas, resulta realmente complicado (cuando no imposible, incluso para programadores expertos), conocer por completo como la máquina reacciona en cada situación, pero sí parece ser exigible (aunque, en mi opinión, la IA deberá precisar este concepto en nuevas regulaciones) haber sometido a la IA a controles suficientes como para garantizar una reacción normal ante situaciones normales.

En el caso de la IA mencionada, parece una situación normal que mujeres y afroamericanos pidan el subsidio, por lo que debería haber sido considerado como parte de la diligencia debida de aquellos programadores de la máquina la comprobación de la inexistencia de sesgos debido únicamente a estas características.

Al no haberse cumplido la exigencia debida, la responsabilidad en este caso correspondió a los programadores y, por ello, a la empresa desarrolladora.

Pero... ¿ha de acudir al caso concreto? ¿y queda la imputación de la responsabilidad en manos del juez? ¿deberían regularse algunas presunciones (y en su caso, se tratarían de presunciones *iuris tantum* o *iuris et de iure*)? ¿y sobre quién recae la carga de la prueba?

- ¿Tienen las IAs personalidad jurídica propia?

Muy unido al tema de la responsabilidad de las IAs encontramos la necesidad de regular la IA a nivel jurídico. Si hemos afirmado que existe la posibilidad de que la IA tenga, en ocasiones, responsabilidad, es imperativo que esta tenga personalidad jurídica propia.

¹¹⁵ <https://www.va.gov/>

¹¹⁶ Máster en IA e Innovación, Founderz y Microsoft. Clase M4/12, Rendición de Cuentas y Personalidad – Introducción, minuto 7:14.

Sin embargo, y cómo se explicó en el apartado correspondiente¹¹⁷, el encaje en el sistema jurídico actual no es sencillo.

Si se reconociera personalidad jurídica propia a la IA se estaría admitiendo que esta puede tomar decisiones, y que debe responder por ellas. Al efecto, ¿hasta qué punto podemos considerar que una IA puede “decidir”? ¿no sería otorgar características humanas a algo inerte? ¿y cuál sería el “castigo” o “pena” por actuación incorrecta?

Parece que la opción más viable sería reconocer a esta una personalidad jurídica electrónica, de acuerdo con lo ya recomendado por la Resolución del Parlamento Europeo de 16 de febrero de 2017. Sin embargo, esta solución genera nuevos inconvenientes. Si se le reconoce personalidad jurídica, ¿se le deben reconocer derechos? y en su caso, ¿cuáles?.

Considero que este complejo tema habrá de ser regulado a nivel supranacional de manera que genere un imperativo homogéneo en todos los países miembros. Es un hecho que las empresas que principalmente han desarrollado (y están desarrollando) IA son empresas globales, por lo que el tratamiento jurídico de las mismas dentro de la UE debería ser, igualmente, uniforme.

Sin embargo, surgen nuevas cuestiones. En casos como el Loomis, ¿debería haber sido tomada como responsable COMPAS por haber realizado un informe inexacto?

En caso de respuesta afirmativa ¿cómo se lleva esta responsabilidad a la práctica?

En caso de que no se entienda responsable a la máquina, sino al fiscal que utilizó dicho informe, ¿ha este de pagar por negligencia? ¿pero cometió realmente negligencia, si utilizó un dictamen pensando que este era correcto? ¿hasta dónde llegan los límites de la diligencia debida?

Y, de manera análoga ¿habría de tomarse responsable al juez por basar su sentencia en las alegaciones del fiscal, sin comprobar de manera oportuna la fiabilidad de sus argumentos? Y en caso afirmativo ¿cuál habría sido la comprobación necesaria?

Si se considera que sí existió la diligencia debida... ¿quién es responsable y cómo se salvaguardan los derechos del perjudicado?

En definitiva, parece indudable que muchos son los desafíos del legislador antes de que la IA pueda, de manera plausible, comenzar a formar parte de nuestro sistema jurídico.

Demasiados cabos sueltos, demasiadas injerencias en nuestros derechos más básicos, para los que todavía no se ha encontrado una vía de solución. Considero que las preguntas arriba expuestas son aquellas que deberán ser analizadas en un primer momento por el legislador europeo y que, solo tras encontrar una solución satisfactoria a todas ellas, podremos comenzar a pensar en el papel de la IA en nuestro entorno jurídico.

Hasta entonces, considero que todo intento de sustitución (e incluso de herramienta con capacidad de decisión propia, todo aquello que vaya más allá de un simple informe sin

¹¹⁷ Me remito a lo explicado en el apartado V. 1. Inteligencia Artificial como sustituto del magistrado. Introducción.

toma de decisiones, y siempre revisado por un humano) es adelantarnos a lo necesario, generando grandes riesgos a cambio de una baja recompensa.

En el ámbito privado la óptica es diversa. Allí, el uso de la tecnología no está colisionando de manera tan frontal con los derechos del ciudadano (aunque sí es cierto que debería tenerse en cuenta el problema de la responsabilidad. ¿qué ocurre si un ciudadano pierde un juicio que debería haber ganado, al haber hecho su abogado un mal uso de la IA?), y son los propios despachos los que, quizá de manera interna, podrán decidir qué uso dar a esta nueva tecnología. Sin embargo, habrá de regularse cómo el uso incorrecto de dicha tecnología en el juicio. ¿deberemos quizá crear un nuevo tipo penal que genere responsabilidad al despacho (o al abogado) por utilizar análisis, referencias o argumentos incorrectos generados por IA? ¿o será la misma responsabilidad que la derivada de haberlos escrito por sí mismo? ¿será responsable el abogado, si ha utilizado de manera diligente las herramientas puestas a su disposición por su despacho? ¿o lo será la firma cómo tal? Y, en su caso, ¿cuál será la responsabilidad exigible? ¿penal? ¿administrativa? ¿disciplinaria?

Y ¿puede hacer algo el ciudadano, que pierde un juicio por un mal uso de la IA por su abogado? ¿deberemos generar un recurso especial? ¿es acaso motivo de recurso que la denegación de la petición se falle en la inexactitud de pruebas generadas por la IA?

Por último, cerraré esta reflexión planteando la cuestión de la idoneidad.

Cómo he analizado con anterioridad, en casos de uso como COMPAS o VioGén, que exista la posibilidad de automatizar no significa que sea, en todo caso, la opción a elegir.

Considero que, hoy en día, existe una especie de fiebre por la automatización. La novedad conlleva una cierta emoción por la implementación de esta tecnología, dejando quizá de lado en ocasiones su utilidad real.

Parece, cayendo en la falacia de la fiabilidad de la máquina, que esta va a hacer un mejor trabajo, siempre y en todo caso, y que por eso su implementación es positiva. Creo que necesitamos unos años para, como sociedad, acostumbrarnos a convivir con esta nueva tecnología. Su desarrollo, en algunos campos, puede ser verdaderamente revolucionario, pero en otros campos los humanos seguiremos siendo la opción a elegir.

Por otro lado, se plantea el dilema moral de si esta va a ser capaz de sustituir los humanos en sus funciones. De nuevo, creo que una vez pasada la fiebre inicial generada por la novedad, y analizado más en profundidad los beneficios, riesgos y posibilidades de su implementación en los diversos sectores (como humildemente he tratado de realizar yo en este trabajo), veremos que los humanos tenemos muchas cualidades que aportar que para una máquina son, simplemente inalcanzables.

Como reflexión final, no, no creo que la IA vaya a sustituir, al menos en las próximas décadas, a jueces y magistrados. Queda demasiado por decidir, y nos enfrentamos a dilemas en su regulación que han de ser pensados lentamente, y que, considero, deben resolverse a nivel Unión Europea (si no a nivel global). En cuanto a su utilización como

herramienta para jueces y magistrados, creo que pueden ser útiles para realizar tareas repetitivas, eminentemente matemáticas, y en las que no tengan que tomar decisiones ni actuar dentro del ámbito de discrecionalidad de la ley. Como Microsoft Excel en su momento revolucionó la contabilidad (pero no eliminó la figura del contable), este tipo de herramientas (como calculadoras de intereses automáticos) pueden ser una ayuda, pero no un sustituto de la labor del juez.

En el ámbito privado, parece que están empezando a desarrollarse sistemas propios que cubren necesidades específicas de los despachos de abogados. Habrá de regularse la cuestión de la responsabilidad, y los diferentes despachos deberán reflexionar acerca de la conveniencia de su adopción en su caso concreto. Sin embargo, mi recomendación para estos sería la paciencia. Pese a que pueda parecer una mejora de productividad, y aunque pueda ser cierto respecto a ciertas tareas de poco valor (como quizá, la búsqueda de jurisprudencia), considero que muchas de estas tareas son necesarias como parte del “entrenamiento” de los jóvenes abogados.

El mayor activo de un despacho de abogados es, sin duda, sus personas, y el conocimiento y la valía de los mismos se desarrollan tras años de ejercicio profesional.

Eliminar la necesidad de juniors, al automatizar las tareas típicamente desarrolladas por estos, podría generar un desequilibrio en la pirámide de crecimiento de estas firmas que lleve a que, de aquí a unos años, no existan personas capacitadas para realizar las tareas que sí generan más valor y que no pueden ser automatizadas.

Sin juniors, no hay seniors.

¿Es un precio que estamos dispuestos a pagar?

IX. Anexo I

Conversación con ChatGPT acerca de su opinión sobre si la Inteligencia Artificial puede sustituir a Jueces y Magistrados.

ChatGPT ▾

¿Crees que la Inteligencia Artificial puede sustituir a los Jueces y Magistrados?



La idea de que la Inteligencia Artificial (IA) pueda sustituir a los jueces y magistrados es un tema complejo y multifacético que merece un análisis detallado. A continuación, presento varios puntos que deben considerarse:

Ventajas Potenciales de la IA en el Sistema Judicial

1. **Eficiencia y Rapidez:** La IA puede procesar grandes cantidades de información de manera rápida y eficiente, lo que podría acelerar los procesos judiciales y reducir la carga de trabajo de los tribunales.
2. **Consistencia y Objetividad:** La IA puede aplicar la ley de manera uniforme y sin sesgos personales, lo que podría conducir a decisiones más consistentes y justas.
3. **Acceso a Información:** La IA puede ayudar a jueces y abogados a acceder a precedentes y leyes relevantes de manera más rápida, mejorando la calidad de las decisiones judiciales.

Desafíos y Limitaciones

1. **Interpretación y Contexto:** La aplicación de la ley no es simplemente un ejercicio mecánico. Requiere una comprensión profunda del contexto, la intención legislativa y las circunstancias individuales de cada caso, algo que la IA todavía no puede manejar completamente.
2. **Sesgos en los Algoritmos:** Los sistemas de IA pueden perpetuar y amplificar sesgos existentes si se entrenan con datos sesgados. La justicia requiere una evaluación cuidadosa y equitativa, algo que podría estar comprometido si los algoritmos no son transparentes y justos.
3. **Ética y Responsabilidad:** Las decisiones judiciales tienen implicaciones profundas para los derechos y libertades individuales. Delegar estas decisiones a una IA plantea serias cuestiones éticas y de responsabilidad, ya que sería difícil asignar responsabilidad por errores judiciales a una máquina.
4. **Empatía y Juicio Humano:** Los jueces humanos pueden mostrar empatía y juicio moral, aspectos cruciales en muchos casos judiciales, especialmente en aquellos que involucran circunstancias excepcionales o mitigan factores humanos.

Rol Complementario de la IA

En lugar de reemplazar completamente a jueces y magistrados, la IA podría desempeñar un rol complementario. Por ejemplo:

- **Asistencia en la Investigación:** Ayudar en la investigación legal, revisión de documentos y análisis de precedentes.
- **Automatización de Tareas Administrativas:** Automatizar tareas rutinarias para permitir a los jueces concentrarse en los aspectos más críticos de los casos.
- **Análisis Predictivo:** Ofrecer análisis predictivos para ayudar a los jueces a comprender posibles consecuencias de sus decisiones basadas en datos históricos.

Conclusión

La IA tiene el potencial de mejorar varios aspectos del sistema judicial, pero su capacidad para sustituir completamente a los jueces y magistrados es limitada por la necesidad de juicio humano, empatía, y comprensión contextual. Por ahora, es más realista y prudente considerar a la IA como una herramienta complementaria que puede asistir y mejorar la eficiencia del sistema judicial, en lugar de un reemplazo total para los jueces humanos.

¿Qué tendría que ocurrir para que la Inteligencia Artificial sea capaz de sustituir por completo a Jueces y Magistrados?



Para que la Inteligencia Artificial (IA) sea capaz de sustituir por completo a jueces y magistrados, se necesitaría un avance significativo en múltiples áreas tecnológicas, legales, éticas y sociales. Aquí se detallan algunos de los desarrollos y condiciones clave que deberían ocurrir:

1. Avances Tecnológicos Significativos

a. Comprensión Contextual y Emocional

- **IA con Capacidad de Interpretación Compleja:** La IA debería ser capaz de comprender no solo el texto legal, sino también el contexto en el que se aplican las leyes, incluyendo las intenciones legislativas y las circunstancias individuales de cada caso.
- **Empatía y Juicio Moral Simulado:** La IA necesitaría desarrollar una forma de simular empatía y juicio moral, para poder considerar factores humanos y emocionales en sus decisiones.

b. Eliminación de Sesgos

- **Algoritmos Transparentes y Justos:** Se deben desarrollar algoritmos de IA que sean completamente transparentes y auditables, con mecanismos robustos para eliminar sesgos y garantizar decisiones equitativas.
- **Datos de Entrenamiento Diversos y Representativos:** Los sistemas de IA deben entrenarse con datos que representen de manera justa y diversa a todas las poblaciones, evitando perpetuar sesgos históricos.

2. Cambios en la Infraestructura Legal y Normativa

a. Adaptación de las Leyes

- **Reformas Legales:** Las leyes y regulaciones deben adaptarse para permitir y regular el uso de IA en el sistema judicial, incluyendo mecanismos de responsabilidad y apelación.
- **Estándares Internacionales:** Se necesitan estándares internacionales para garantizar que la implementación de la IA en el sistema judicial sea coherente y justa a nivel global.

b. Supervisión y Gobernanza

- **Órganos de Supervisión:** Crear organismos de supervisión independientes para monitorear y evaluar el desempeño de la IA en el sistema judicial, asegurando su responsabilidad y transparencia.
- **Mecanismos de Apelación:** Establecer procedimientos claros para apelar decisiones tomadas por IA, permitiendo la intervención humana en casos excepcionales.

3. Desarrollo de Confianza Pública

a. Transparencia y Educación

- **Transparencia en el Funcionamiento de la IA:** Las decisiones de la IA deben ser transparentes y explicables, permitiendo a las partes involucradas comprender cómo se llegó a una determinada decisión.
- **Educación Pública:** Informar y educar al público sobre cómo funciona la IA en el sistema judicial y sus beneficios y limitaciones.

b. Pruebas y Validaciones Extensivas

- **Proyectos Piloto y Pruebas Extensas:** Implementar proyectos piloto y pruebas exhaustivas para validar la eficacia y justicia de la IA en diversos contextos judiciales antes de una adopción amplia.
- **Evaluación Continua:** Realizar evaluaciones continuas para ajustar y mejorar los sistemas de IA basados en retroalimentación y resultados reales.

4. Resolución de Cuestiones Éticas

a. Consideraciones Éticas y Morales

- **Debates Éticos Amplios:** Iniciar debates éticos y filosóficos amplios sobre el uso de IA en decisiones judiciales, considerando las implicaciones morales de delegar tales responsabilidades a máquinas.
- **Consenso Social:** Alcanzar un consenso social sobre el papel de la IA en el sistema judicial, asegurando que su uso refleje los valores y principios de la sociedad.

Conclusión

La sustitución completa de jueces y magistrados por IA requeriría una transformación integral y profunda en múltiples aspectos tecnológicos, legales, éticos y sociales. Aunque la IA puede complementar y mejorar varios aspectos del sistema judicial, lograr una sustitución total es un desafío monumental que, en el estado actual de la tecnología y la sociedad, parece lejano. La evolución hacia un sistema judicial altamente asistido por IA debe hacerse con cautela, asegurando siempre la justicia, equidad y respeto por los derechos humanos.



X. Bibliografía

- Castellano, P. S. (2021). Inteligencia Artificial y Administración de Justicia: ¿Quo vadis, justitia? *Revista de Internet, Derecho y Política*.
- Colomer, J.-L. G. (s.f.). Sobre una posible y futura responsabilidad civil del juez-robot: una aproximación desde el derecho procesal. *Derecho Digital*.
- Díaz, M. C. (2022). Garantías Constitucionales y Jurisdiccionales antes el desarrollo de herramientas de Jurimetría Analítica. Especial consideración a la tutela antidiscriminatoria ante la llamada "justicia predictiva". ¿Dónde estamos? *Huri-Age Red Tiempo de los Derechos* n° 5.
- España, G. d. (2020). El Libro blanco de la Inteligencia Artificial aboga por la excelencia y la confianza. *Datos.gob.es*.
- Fenoll, J. N. (2023). The loss of digital control: are we heading towards a judicial dystopia? *Smarteca, Actualidad Civil*, n° 4.
- Fouquet, D. B. (2022). Sobre Inteligencia Artificial. Decisiones Judiciales y Vacíos de Argumentación.
- IBM. (2023). ¿Qué es el sesgo de la IA? Obtenido de <https://www.ibm.com/es-es/topics/ai-bias#:~:text=El%20sesgo%20de%20IA%2C%20tambi%C3%A9n,resultados%20distorcionados%20y%20potencialmente%20perjudiciales>.
- Manuel Olmedo, S. G. (2023). El uso de la IA puede reducir el tiempo dedicado a la atención de casos y procesos judiciales, permitiendo a los profesionales de la Justicia abordar más expedientes y acortar los procesos. *Diariolaley*.
- Matín, N. B. (2022). Algoritmos predictivos al servicio de la Justicia: ¿Una nueva forma de minimizar el riesgo y la incertidumbre? *Revista da Faculdade Mineira de Direito* n° 43.
- Navarro, S. N. (s.f.). Responsabilidad civil e Inteligencia Artificial. *El Cronista del Estado Social y Democrático de derecho*.
- Notaro, L. (s.f.). "Algoritmos predictivos" y justicia penal desde una perspectiva italiana y europea. En L. Notaro.
- Ortega, R. R. (s.f.). ¿Pueden los robots remplazar a los funcionarios? *Diariolaley*.
- Pérez, C. L. (2022). *Inteligencia Artificial en la Administración de Justicia: Regulación española y marco europeo e internacional. Proyectos desarrollados por el Ministerio de Justicia de España*. Centro de Estudios Jurídicos.
- Pulido, M. d. (2023). Ética de la Inteligencia Artificial jurídica aplicada al proceso. *Cuadernos Electrónicos de Filosofía del Derecho*.
- Román, J. C. (2021). LA evolución hasta los servicios no presenciales en la Administración de Justicia. *Centro de Estudios Jurídicos*.
- Sancho, M. d. (2021). El Libro Blanco sobre Inteligencia Artificial de la Comisión Europea: Reflexiones desde las Garantías esenciales del proceso penal como "sector de riesgo". *Revista Española de Derecho Europeo* Núm. 76.

- Sancho, M. d. (2021). El uso Jurisdiccional de los Sistemas de Inteligencia Artificial y la necesidad de su armonización en el contexto de la Unión Europea. *Revista General de Derecho Procesal* 55.
- Velasco, M. J. (2023). Uso ético de Inteligencia Artificial en Justicia. *Diariolaley*.
- Vilar, S. B. (2023). Dataización de la Justicia (Algoritmos, Inteligencia Artificial y Justicia, ¿El comienzo de una gran amistad?). En *Revista Boliviana de Derecho* n°36 (págs. 14-45).
- Vilar, S. B. (s.f.). La seductora algoritmización de la justicia. Hacia una justicia poshumanista (Justicia+) ¿utópica o distópica? *El cronista del Estado Social y Democrático de Derecho*.