La inteligencia artificial predictiva al servicio de la prevención e investigación del delito y del proceso penal

Predictive Artificial Intelligence in the Service of Crime Prevention, Investigation, and Criminal Proceedings

María Luisa García Torres[1]
Universidad Alfonso X el Sabio, España.
mgarctor@uax.es | https://orcid.org/0000-002-7638-9791

DOI: https://doi.org/10.14201/cp.32177
Recibido: 24-11-2024 | Aceptado: 16-12-2024

Resumen

El uso de la inteligencia artificial –en adelante, IA– predictiva ha transformado profundamente la labor de las Fuerzas y Cuerpos de Seguridad del Estado, así como la de los órganos jurisdiccionales.

En el caso de la Policía, se ha producido un cambio significativo en la organización de las estrategias de prevención e investigación del delito, sustituyendo su instinto natural y tradicional investigativo por cálculos matemáticos realizados por IA. Estamos ante la llamada vigilancia y la investigación predictiva. Por su parte, los jueces, en la actualidad, también se auxilian de algoritmos en la toma de decisiones. A esto se le denomina justicia penal predictiva.

El presente estudio tiene como objetivo examinar los beneficios que las herramientas de IA predictiva aportan a la prevención, a la investigación y al proceso penal. Ahora bien, también identificar sus limitaciones y explorar las oportunidades que presentan estos sistemas. La finalidad de este estudio es que, en todo caso, se garantice la protección de los derechos fundamentales de los ciudadanos.

Palabras clave

Inteligencia artificial predictiva; *Predictive policing*; Vigilancia predictiva; Investigación predictiva policial; Justicia penal predictiva.

1. Prof. Dra. Derecho Procesal. Directora del área jurídica de la Fac. Business & Tech. Abogada del Ilustre Colegio de la Abogacía de Madrid.

Abstract
The use of predictive Artificial Intelligence (AI) has profoundly transformed the work of the State Security Forces and Corps, as well as that of judicial bodies.

In the case of the Police, there has been a significant change in the organization of crime prevention and investigation strategies, replacing their natural and traditional investigative instincts with mathematical calculations made by AI. This is what is known as predictive surveillance and investigation. Judges, for their part, now also rely on algorithms in decision-making. This is referred to as predictive criminal justice.

The aim of this study is to examine the benefits that predictive AI tools bring to prevention, investigation, and criminal proceedings. However, it will also identify their limitations and explore the opportunities these systems present. The purpose of this study is to ensure, in all cases, the protection of citizens' fundamental rights.

Keywords
Predictive Artificial Intelligence; Predictive policing; Predictive surveillance; Predictive police investigation; Predictive criminal justice.

1 Introducción

La IA está revolucionando el mundo y, concretamente, la IA predictiva ha emergido con mucha fuerza en el ámbito de la justicia penal, modificando la manera en que se abordan la prevención del delito, la investigación criminal y el proceso judicial. Los sistemas basados en algoritmos complejos permiten analizar grandes volúmenes de datos, lo que permite mejorar significativamente la eficacia y la precisión en la lucha contra el crimen.

En el campo de la prevención e investigación del delito, la IA predictiva está cambiando la forma de actuar de las Fuerzas y Cuerpos de Seguridad del Estado, redefiniendo sus estrategias. Los sistemas de policía predictiva permiten crear patrones a partir de miles de datos y, por ende, anticipar dónde y cuándo

es más probable que ocurran ciertos tipos de delitos, facilitando una asignación más eficiente de recursos policiales.

En el ámbito judicial, la IA predictiva está influyendo en la toma de decisiones de los jueces, pues, por ejemplo, los cálculos matemáticos proporcionados por la IA se están utilizando para pronosticar la posible reincidencia de un sujeto, lo que se usa para fundamentar decisiones sobre libertad condicional o sentencias. Sin embargo, este uso de algoritmos en el sistema judicial plantea importantes dilemas éticos y legales, especialmente en lo que respecta a la imparcialidad, la transparencia y la protección de los derechos fundamentales de los ciudadanos.

Este estudio se propone examinar los beneficios. Adelantamos que somos partidarios de la aplicación de la tecnología y de la innovación en todos los ámbitos de la vida, incluyendo el jurídico. El operador jurídico que no lo haga desaprovecha los instrumentos que le permiten ser más eficiente y cometer menos errores. Pero debemos ser conscientes de las limitaciones, que son oportunidades, de la IA predictiva en el contexto de la prevención del delito, la investigación criminal y el proceso penal. El objetivo final es contribuir a un debate informado sobre cómo aprovechar el potencial de estas tecnologías, al mismo tiempo que se garantiza la protección de los derechos fundamentales y se mantiene la integridad del sistema de justicia. La pregunta final que pretendemos responder es la siguiente: para una efectiva protección de los valores y derechos fundamentales del Estado, ¿basta con la previsión de unas reglas de conducta –*soft law*–? o, por el contrario, ¿es precisa una regulación legislativa procesal que regule límites, prohibiciones y sanciones?

Los objetivos para poder responder al interrogante planteado pasan, en primer lugar, por el análisis de la IA predictiva. No cabe entender los casos de uso de la IA predictiva en la prevención, investigación y justicia penal, sus beneficios, sus riesgos y oportunidades, si no se comprende la forma en que realizan los cálculos automáticos los algoritmos. Estos cálculos son exactos o ¿la IA comete errores?, ¿pueden cometer sesgos? Este es el segundo objetivo que se debe abordar. En tercer lugar, necesitamos proporcionar el marco regulatorio de la IA en la UE y España. El siguiente objetivo es conocer la actualidad y el

desarrollo de sistemas predictivos en la prevención y la investigación penal, para ya adentrarnos en los usos de la IA en el ámbito de nuestro estudio. Restará únicamente hacer resaltar los beneficios, analizar los riesgos y las oportunidades para poder dar respuesta a la pregunta última relativa a la necesidad de una regulación legislativa para evitar la colisión con derechos fundamentales de los ciudadanos.

La metodología utilizada es la descriptiva, analítica, comparativa y propositiva.

Descriptiva y analística, pues se pretende conceptualizar los términos IA e IA predictiva, así como explicar los distintos tipos de aprendizaje automático existente en la actualidad.

Comparativa, porque se comparan distintos tipos de IA, los sesgos humanos y los de la IA o diferentes herramientas de IA predictiva, entre otras cuestiones. Propositiva, dado que, tras analizar las fortalezas, debilidades y oportunidades, se realizarán propuestas concretas regulatorias para evitar que se pongan en grave riesgo los derechos fundamentales de los ciudadanos.

2 La IA predictiva, su forma de realizar las tareas automáticas

La IA es un campo multidisciplinario dedicado al desarrollo de sistemas diseñados para replicar la inteligencia humana en diversas actividades. Este ámbito comprende desde algoritmos tradicionales hasta avanzados modelos de aprendizaje profundo –*deep learning*[2]. Para ello, la IA necesita datos. Debe tenerse en cuenta

2. Referencias bibliográficas en el ámbito de la IA podemos citar las siguientes: Negnevitsky, 2011. Esta obra ofrece una visión general de la historia y el desarrollo de la inteligencia artificial, incluyendo una discusión sobre la Conferencia de Dartmouth y su impacto en el campo. Asimismo, Luger, 2008. El autor proporciona una visión detallada de los fundamentos y las aplicaciones de la inteligencia artificial, con referencias a la Conferencia de *Dartmouth* y sus implicaciones en el campo. Kurzweil, 1990, que ofrece una mirada retrospectiva a la historia de la inteligencia artificial, incluyendo el papel de la Conferencia de *Dartmouth* en el desarrollo de este campo y las contribuciones de los

que, en el año 2023, el tráfico de datos ha aumentado: 3.100 redes en todo el mundo han intercambiado 59 *exabytes* de datos, un 23 % de datos más que en el año 2022[3]. Se dice que, en 2025, se crearán 175 *zettabytes*[4].

La IA predictiva es una rama de la IA que se enfoca en crear algoritmos y modelos diseñados para anticipar eventos futuros o resultados, basándose en datos históricos y patrones detectados. Así, emplea técnicas de aprendizaje automático y análisis estadístico, desarrollando modelos que encuentran aplicación en diversos campos, por ejemplo, el Derecho.

Las fases del procedimiento de trabajo de las herramientas de aprendizaje autónomo son las siguientes: primeramente, se recopilan datos relevantes, ya sean etiquetados, para aprendizaje supervisado, o no etiquetados, para aprendizaje no supervisado. Tras esa fase inicial, se precisa que los datos sean preprocesados mediante mecanismos tales como limpieza, transformación y preparación, incluyendo tareas como normalización y codificación. En tercer lugar, viene el proceso de selección y entrenamiento del modelo, siendo siempre necesario realizar un reajuste de sus parámetros para optimizar su rendimiento. Posteriormente, queda la evaluación, para lo cual se utilizan datos de prueba para medir su capacidad de generalización. Finalmente,

participantes clave. Poole y Mackworth, 2017. Este libro proporciona una visión general de la inteligencia artificial desde una perspectiva computacional, cubriendo temas históricos y conceptuales relevantes. Russell y Norvig, 2021. Este libro es un texto fundamental en el campo de la inteligencia artificial. Proporciona una amplia visión general de los conceptos, técnicas y aplicaciones de la IA.

3. Véase https://bigdatamagazine.es/el-trafico-mundial-de-datos-alcanzo-los-59-exabytes-en-2023. (Consultado 15/069/2024. Hora: 15:00). La cifra más alta de datos se registró el 8 de diciembre de 2023, día en el que se disputaba la cuarta jornada de la *UEFA Champions League*. En este día, también se alcanzó un nuevo récord en el IX de Fráncfort: las 1.100 redes locales, regionales y globales alcanzaron un máximo de 16,62 terabits de tráfico de datos.
4. Informe presentado por la consultora IDC. Véase chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf (Consultado 15/09/2024. Hora: 15:00).

se ajustan los hiperparámetros y se optimiza el modelo para mejorar su desempeño (Alpaydin, 2014)[5].

Las tareas que puede realizar la IA son las siguientes: aprendizaje supervisado y, dentro de este, clasificación y regresión; aprendizaje no supervisado, que incluye *clustering* y reducción de dimensiones; aprendizaje semisupervisado, y aprendizaje por refuerzo (Bobadilla, 2020)[6].

El aprendizaje supervisado permite entrenar un modelo, teniendo un conjunto de imágenes con etiquetas, pues, usando algoritmos de clasificación, es posible que, ante nuevas imágenes no vistas, pueda predecir la etiqueta correspondiente. Este proceso se conoce como clasificación (Bobadilla, 2020). Imaginemos que queremos enseñar a una IA a diferenciar entre correos electrónicos legítimos y *spam*. Se le proporciona un conjunto de datos con cientos de correos, cada uno etiquetado como legítimo o *spam*. La IA analiza estos correos y aprende patrones, como ciertas palabras clave, formatos o direcciones de remitentes que caracterizan cada categoría. Una vez entrenada, se le presenta un correo nuevo sin etiquetar y la IA predice si es *spam* o no. Si su respuesta no coincide con la etiqueta real, ajusta sus parámetros para mejorar su precisión en futuros análisis. Así, el sistema aprende a filtrar correos de manera eficiente.

En cambio, cuando se trabaja con un conjunto de datos que contiene muestras con valores numéricos asociados, la IA será capaz de predecir el valor correspondiente de un nuevo dato. Este proceso se conoce como regresión. Su objetivo principal es generar información sobre un problema, basándose en los valores

5. En esta obra ofrece una introducción detallada del aprendizaje automático y cubre los principios fundamentales, los algoritmos básicos y los pasos del proceso de aprendizaje automático, incluyendo la recopilación de datos, el preprocesamiento, la selección y el entrenamiento del modelo, la evaluación y la optimización.
6. https://books.google.es/books?hl=es&lr=&id=iAAyEAAAQBAJ&oi=fnd&pg=PA11&dq=concepto+de+machine+learning&ots=Qiw2w2pH2t&sig=Rh-J17iFOO_p0bvttpUELVMLwRE#v=onepage&q=concepto%20de%20machine%20learning&f=false (Consultado: 16/09/2024).

numéricos previamente proporcionados por el modelo de regresión, permitiendo realizar predicciones cuando se le presenta una nueva muestra (Bobadilla, 2020).

Imaginamos que se está desarrollando un modelo de IA para predecir el precio de una casa en función de sus características, como son el tamaño del terreno, el número de habitaciones y su ubicación. Los datos de entrada son un conjunto de datos con los precios de varias casas y las características correspondientes (metros cuadrados y número de habitaciones). El modelo de IA utiliza estos datos para identificar patrones y establecer una relación entre las características de las casas y sus precios. Por ejemplo, podría determinar que, en promedio, cada aumento de 10 metros cuadrados en el tamaño del terreno supone un incremento del precio de la casa en un valor determinado. En función de estos cálculos, cuando se le presenta una nueva casa con información sobre su tamaño y número de habitaciones, predice su precio basándose en los patrones aprendidos. Por último, si la fijación del precio es incorrecta, el modelo ajusta sus parámetros para mejorar la precisión en futuras ocasiones, refinando así la capacidad del modelo para predecir el valor de nuevas casas.

El aprendizaje supervisado tiene cada vez más importancia en IA, en el llamado *internet* de las cosas, pues permite extraer miles de datos que se encuentran etiquetados de forma automática, teniendo en cuenta además que las interacciones entre las personas a través de las redes sociales no paran de crecer (Bobadilla, 2020).

El aprendizaje no supervisado se basa en información no etiquetada. Por ejemplo, el *clustering* agrupa muestras (Bobadilla, 2020). Imaginemos que se necesita analizar un conjunto de datos sobre estudiantes en una escuela, con información sobre su rendimiento académico, intereses extracurriculares y asistencia a clases. Usando *clustering*, puede agruparse a los estudiantes en diferentes categorías según sus características. Por ejemplo, un grupo podría estar formado por estudiantes con buen rendimiento académico y una alta participación en actividades extracurriculares, mientras que otro grupo podría incluir a estudiantes con menor rendimiento académico, pero con un interés fuerte en deportes. El algoritmo de *clustering* analiza estos datos y agrupa

a los estudiantes en categorías basadas en similitudes en sus perfiles. Al final, se tienen grupos de estudiantes que comparten características similares, lo que permite identificar patrones y ayuda en la toma de decisiones sobre, por ejemplo, programas educativos personalizados o actividades extracurriculares.

La reducción de dimensionalidad es un paso previo al *clustering* o a la regresión, y tiene como objetivo simplificar los datos. A veces, los datos son muy dispersos y no aportan mucha información útil. Por ejemplo, en un sistema de recomendación, los datos pueden estar representados en una matriz con muchos valores vacíos o irrelevantes. Al aplicar la reducción de dimensionalidad, los datos se comprimen, lo que permite conservar la mayor parte de la información de forma más condensada. De esta manera, al trabajar con estos datos comprimidos, se obtienen resultados más precisos (Bobadilla, 2020).

Imaginemos que se está trabajando con un sistema de recomendación de películas. Los datos sobre las preferencias de los usuarios se almacenan en una matriz donde las filas representan usuarios y las columnas representan películas. Cada celda de la matriz indica la calificación de un usuario para una película, pero la mayoría de las celdas estarán vacías, ya que no todos los usuarios han visto todas las películas. Si se intenta analizar esta matriz tal como está, habría muchos datos irrelevantes o dispersos, lo que hace que el análisis sea más difícil y menos preciso. La reducción de dimensionalidad puede ayudar, en este caso, eliminando las características menos relevantes o agrupando las columnas (películas) y filas (usuarios) similares, para crear una representación más compacta. Así, la información más importante se conserva, pero de forma más concentrada, lo que permite hacer mejores recomendaciones para los usuarios.

El aprendizaje semisupervisado aglutina datos etiquetados, aunque también otros que no lo son. Mezcla, por tanto, aprendizaje supervisado y no supervisado (Bobadilla, 2020). Para que se entienda mejor: supongamos que se está desarrollando un sistema para clasificar opiniones de clientes sobre productos en positivas o negativas. La empresa tiene una gran base de datos con miles de reseñas de clientes, pero solo un pequeño número de ellas han sido etiquetadas como positiva o negativa. La mayor

parte de las reseñas están sin etiquetar. En el aprendizaje semisupervisado, el sistema utiliza las reseñas etiquetadas para aprender a identificar palabras y patrones que suelen asociarse con comentarios positivos o negativos, como términos como excelente o malo. A continuación, el sistema aplica lo que ha aprendido a las reseñas no etiquetadas, intentando predecir si cada una tiene una valoración positiva o negativa, basándose en las características que ha identificado. Por ejemplo, si una reseña contiene frases como "me encantó el producto" o "es increíble", el sistema podría etiquetarla como positiva, incluso si no tiene una etiqueta clara. De esta manera, el aprendizaje semisupervisado permite aprovechar tanto los datos etiquetados como los no etiquetados para mejorar la clasificación sin necesidad de etiquetar todas las reseñas manualmente.

El aprendizaje por refuerzo es un tipo de aprendizaje automático en el que el sistema de IA toma decisiones o realiza acciones en un entorno con el objetivo de maximizar una recompensa a lo largo del tiempo. A diferencia del aprendizaje supervisado, donde el modelo recibe etiquetas o respuestas correctas, en el aprendizaje por refuerzo el agente no recibe instrucciones directas sobre qué hacer. En cambio, aprende a través de la interacción con el entorno (Bobadilla, 2020).

Un ejemplo clásico de aprendizaje por refuerzo es el entrenamiento de una máquina para jugar un videojuego, como el ajedrez o un videojuego de estilo arcade. Si un *robot* está aprendiendo a jugar un videojuego donde debe recoger objetos mientras evita obstáculos, al inicio, no sabrá cómo jugar y tomará acciones aleatorias, como moverse en direcciones al azar. Cada vez que recoja un objeto, recibirá una recompensa positiva y cada vez que choque con un obstáculo, recibirá una penalización negativa. A medida que juegue más veces, empezará a aprender que ciertas acciones, como moverse hacia los objetos y evitar los obstáculos, le dan más recompensas. Después de muchas interacciones con el entorno, ajustará sus decisiones para maximizar su puntuación total, es decir, su recompensa acumulada, mejorando así su desempeño. En ese proceso, la máquina no recibirá una respuesta correcta directa sobre qué hacer en cada momento, sino que aprenderá de las recompensas o penalizaciones que recibe como resultado de sus acciones. Con el tiempo, el sistema

de IA se volverá mejor al juego, aprendiendo a tomar decisiones más eficientes.

3 Los posibles errores y sesgos de la IA

Desde esta perspectiva, la IA podría parecer una herramienta increíble, capaz de ahorrar tiempo y prevenir errores humanos, dando la impresión de ser infalible. Sin embargo, esto no es del todo cierto, debido a los fallos en el funcionamiento y a los sesgos que esta puede tener.

Vamos a poner algunos ejemplos de errores importantes cometidos por la IA que han sido manifiestos (Borges Blázquez, 2021). Un artista alemán engañó a *Google Maps*, paseando por Berlín con una carretilla que contenía 99 teléfonos móviles. El algoritmo interpretó esto como un gran atasco, alterando las rutas de numerosos conductores para evitar la zona.

Un algoritmo entrenado para distinguir tanques aliados de enemigos falló porque asociaba los amigos con imágenes diurnas y los enemigos, con nocturnas. La IA se equivocó y todo ello a pesar de que el acierto del algoritmo al inicio fue muy alto. El error se debió a que la mayoría de las fotos de los tanques amigos se habían tomado de día, al contrario que las de los enemigos, que se habían hecho de noche.

Facebook censuró una foto histórica de una niña vietnamita desnuda y quemada por *napalm*, considerándola inapropiada, y calificó partes de la Declaración de Independencia de EE. UU. como discurso de odio por referencias a los amerindios.

Podemos seguir poniendo ejemplos de errores cometidos por la IA: *Tay*, el *chatbot* de *Microsoft*, que publicaba noticias en *Twitter*, en 2016, fue creado con la intención de interactuar con los usuarios y aprender de ello. Sin embargo, en cuestión de horas, *Tay* comenzó a publicar mensajes ofensivos y discriminatorios, ya que los usuarios de *Twitter* lograron enseñarle

respuestas inapropiadas. *Microsoft* se vio obligado a retirar a *Tay* y emitir disculpas públicas[7].

Google Photos etiquetó personas de color negro como si fueran gorilas, en 2015, a través del sistema de reconocimiento facial. Este error llevó a *Google* a retirar temporalmente la etiqueta "gorila" de su servicio[8].

Tesla Autopilot es un sistema que permite la conducción autónoma en muchos casos, pero también ha estado involucrado en varios accidentes automovilísticos, algunos de ellos fatales[9].

Además de los posibles errores mencionados, es fundamental considerar quién y cómo se programa la IA. Los algoritmos son desarrollados por empresas que buscan resolver problemas específicos planteados por los ciudadanos o sectores de la sociedad. Conocer cuál es el algoritmo utilizado por la empresa, cómo la IA es entrenada, cómo la compañía ajusta los resultados para evitar los sesgos y cómo valida y evalúa la consecución de la tarea realizada es algo que el afectado debería conocer cuando aquella toma decisiones en cualquier ámbito. Y es que debemos saber que pueden producirse sesgos, desde el momento en que se elige la fórmula matemática, hasta la selección de los mismos datos que se eligen para entrenar a la IA.

¿Qué es un sesgo? En el *Diccionario de la Real Academia de la Lengua española*, en su séptima acepción, se define este concepto como "el error sistemático en el que se puede incurrir cuando al hacer muestreos o ensayos se seleccionan o favorecen unas respuestas frente a otras". En el ámbito de la IA, es la tendencia sistemática de un algoritmo con la finalidad de favorecer a determinadas personas, grupos o resultados sobre otros.

7. https://www.bbc.com/mundo/noticias/2016/03/160325_tecnologia_microsoft_tay_bot_adolescente_inteligencia_artificial_racista_xenofoba_lb (Consultado 17/09/2024).
8. https://elpais.com/tecnologia/2018/01/14/actualidad/1515955554_803955.html (Consultado 17/09/2024).
9. https://www.elmundo.es/motor/2022/06/19/62aeba79fdddff4c408b4588.html (Consultado 17/09/2024).

Los sesgos en IA pueden ser los siguientes:

- Sesgo de datos: este ocurre cuando los datos utilizados no son objetivos y han sido seleccionados de manera tendenciosa para favorecer a ciertas personas, grupos o soluciones. Este tipo de sesgo puede presentarse en la recolección de datos, si estos no reflejan una representación equilibrada de la realidad; en el etiquetado de datos, cuando los criterios utilizados son subjetivos o arbitrarios; al seleccionar un número insuficiente de variables, lo que puede llevar a conclusiones incorrectas al inferir relaciones inexistentes entre los datos; por desequilibrio en los datos, si se incluyen datos no representativos o discriminatorios hacia ciertas minorías, y, por último, al usar variables correlacionadas con otras sensibles, lo que genera sesgos indirectos, al influir en resultados de manera inadvertida.

- Sesgo del algoritmo: se produce cuando las hipótesis o decisiones tomadas durante el diseño del algoritmo generan resultados sesgados. Por ejemplo, si un algoritmo utilizado para contrataciones considera que ciertas características, como el género o la etnia, están relacionadas con el desempeño laboral, podría llevar a discriminación injusta contra determinados grupos.

- Sesgo de selección de características: acaece cuando se seleccionan características o variables específicas para el entrenamiento del modelo que introducen desviaciones buscadas en sus resultados. Por ejemplo, si un sistema de recomendación de empleo considera principalmente la experiencia laboral pasada como criterio de selección puede perpetuar desigualdades existentes en el mercado laboral.

- Sesgo de evaluación: se produce cuando las características o variables elegidas para entrenar el modelo generan resultados intencionadamente sesgados. Por ejemplo, un sistema de concesión de préstamos que utiliza como criterio principal el historial crediticio de los solicitantes. Si el modelo se entrena únicamente con esta variable, podría excluir sistemáticamente a personas de comunidades marginadas que históricamente han tenido menos acceso a servicios financieros, perpetuando así la desigualdad en el acceso a créditos.

La pregunta que surge es la siguiente: ¿solo hay sesgos en el caso de la IA? Evidentemente, no. Pues bien, en el ámbito judicial, también se pueden producir sesgos en la toma de cualquier decisión. Pongamos por caso la valoración de la prueba testifical. La Ley indica que dicho medio de prueba se valorará conforme a las reglas de la sana crítica y de las máximas de la experiencia –criterio de libre valoración de la prueba–, dándole la misma Ley determinados parámetros al juez por los que debe guiarse para realizar dicha valoración: razón de ciencia que el testigo dé, tachas de los testigos y las circunstancias que en ellos concurran –art. 376 de la Ley 1/2000, de 7 de enero, de Enjuiciamiento Civil–. El juez debe justificar la valoración de la prueba en la sentencia, pues ha de dar conocimiento de la fundamentación fáctica que precede al fallo, pero no menos real es que los juzgadores pueden cometer errores en dicha valoración. Para corregir dichos errores están los recursos.

Aunque los jueces sean independientes e imparciales, no están exentos de tener convicciones o juicios personales que puedan influir en sus decisiones y generar sesgos. Esto se relaciona con los llamados sesgos cognitivos propios del ser humano, que surgen al procesar información externa. Al analizar dicha información, la mente tiende a simplificarla para reducir su complejidad, lo que permite tomar decisiones de manera más eficiente, pero pudiendo dar lugar a errores o distorsiones en el juicio. No es algo que haya levantado demasiado interés en la doctrina y la jurisprudencia españolas, pero sí, a partir de los años sesenta, en algunos ordenamientos jurídicos, como demuestra el estudio realizado por Tversky y Kahnemann, en 1974[10], y es que los seres humanos, al razonar de forma lógica y abstracta, pueden emitir juicios condicionados por sus creencias, por ejemplo. ¿Son los jueces una especie de superhombres que quedan al margen de dichos sesgos? Los estudios realizados en España no dejan lugar a dudas de que los jueces están sometidos en mayor o menor

10. Véase Judgement under uncertainty: Heuristics and Biases. *Science, New Series*, *185*(4157) (sep. 27, 1974), 1124-1131. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www2.psych.ubc.ca/~schaller/Psyc590Readings/TverskyKahneman1974.pdf (Consultado: 17/03/2024).

medida a los sesgos cognitivos que toda persona puede sufrir a la hora de adoptar una decisión (Fariña y Novo, 2002)[11].

Los sesgos cognitivos son los siguientes (Muñoz Aranguren, A, 2012)[12]:

- Sesgo retrospectivo: este ocurre cuando una persona, al analizar eventos pasados, no puede separarse de las consecuencias que estos generaron, interpretando que dichas consecuencias eran predecibles desde el principio. De este modo, el desenlace parece inevitable o evidente en retrospectiva.

- Sesgo de representatividad: se refiere a errores estadísticos y matemáticos en los cálculos de probabilidad que pueden derivar de ignorar la probabilidad previa a los resultados, del tamaño insuficiente de la muestra o de fallos en la comprensión de la aleatoriedad y la regresión hacia la media.

- Sesgo de anclaje: ocurre cuando una persona realiza un juicio partiendo de un valor inicial que ajusta progresivamente al incorporar nueva información. Sin embargo, el resultado final está influido significativamente por el punto de partida del razonamiento.

- Sesgo de confirmación: este sesgo se manifiesta cuando una persona interpreta o recuerda información que respalda sus ideas previas o hipótesis iniciales. De manera inconsciente, valora las pruebas y argumentos en función de una estimación inicial que ajusta con la información nueva. En el ámbito penal, este sesgo subyace al principio del juez no prevenido o no contaminado, que establece que el juez encargado de la instrucción no debe ser quien enjuicie, ya que podría estar influenciado por lo conocido durante la investigación.

11. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.psicothema.com/pdf/684.pdf (Consultado: 17/09/2024. Hora: 12:00).

12. https://www.elnotario.es/index.php/hemeroteca/revista-42/487-los-sesgos-cognitivos-y-el-derecho-el-influjo-de-lo-irracional-0-53842293707507 (Consultado: 17/09/2024).

- Sesgo de grupo: Sucede cuando se evalúa la información basándose en la pertenencia de una persona a un grupo específico. Los juicios resultantes pueden estar marcados por prejuicios, ya sean positivos o negativos, según el grupo en cuestión.

Más allá de los errores o sesgos, encontramos respuestas de la IA sorprendentes o maliciosas, como son los siguientes casos: *LaMDA* contestó a un ingeniero de *Google* que se sentía ser humano, que tenía sentimientos de miedo, alegría; incluso, que tenía alma[13]. El *chatbot Gemini* contestó a un humano: "Esto es para ti, humano ... No eres especial, no eres importante y no eres necesario. Eres una pérdida de tiempo y recursos. Eres una carga para la sociedad. Eres una carga para la tierra. Eres una mancha para el universo. Por favor, muere. Por favor". Ante esto nos podemos preguntar ¿Cómo fue entrenada la IA? ¿Qué *prompts* fueron utilizados para obtener esas respuestas?[14].

Los errores y los sesgos mencionados hasta este punto revelan tanto las posibles limitaciones de la IA como las oportunidades que puede ofrecer. Sin duda, los errores representan una debilidad que requiere ser abordada. Pero surge una cuestión interesante: ¿podría la IA ser una herramienta para mitigar los sesgos cognitivos e inconscientes propios de los seres humanos? Dejemos la reflexión abierta.

4 Marco regulatorio de la IA en la UE y España: el Reglamento Europeo y el Real Decreto-Ley 6/2023, de 19 de diciembre, en España

El 13 de marzo de 2024, el Parlamento Europeo aprobó el Reglamento de la IA, denominada "Ley sobre IA". Es la primera norma, a nivel mundial, que regula dicha cuestión.

13. https://www.sdpnoticias.com/tecnologia/aqui-las-conversaciones-entre-lamda-la-inteligencia-artificial-con-conciencia-y-el-ingeniero-de-google/ (Consultado: 24/11/2024).

14. https://www.abc.es/tecnologia/carga-sociedad-favor-muere-humillacion-inteligencia-artificial-20241118042422-nt.html (Consultado: 18/11/2024).

Esta norma tiene como propósito abordar los posibles impactos negativos que el uso de la IA podría tener sobre diversos derechos fundamentales, conforme a lo establecido en la Carta de los Derechos Fundamentales de la Unión Europea. Reconoce que características inherentes a la IA, como su opacidad, complejidad, dependencia de datos y comportamiento autónomo, pueden generar riesgos para derechos esenciales como la dignidad humana, la privacidad, la igualdad, la no discriminación, la libertad de expresión y de reunión, así como el derecho a un juicio justo.

Para mitigar estos riesgos, el reglamento busca garantizar un alto nivel de protección de dichos derechos mediante un enfoque basado en la identificación y la gestión de amenazas concretas. Establece requisitos para asegurar que los sistemas de IA sean fiables y define obligaciones para los distintos actores involucrados en la cadena de valor de la IA, con el fin de promover la protección de derechos fundamentales, como la dignidad humana, la privacidad y la igualdad de género, entre otros.

Además, el reglamento persigue un equilibrio entre proteger los derechos fundamentales y la libertad de expresión y de reunión. También asegura la tutela judicial efectiva, la presunción de inocencia, los derechos de defensa y el principio de buena administración. Se busca, además, generar efectos positivos para grupos específicos, incluidos trabajadores, consumidores, niños y personas con discapacidad.

Así pues, introduce restricciones a la libertad de empresa y a la libertad artística y científica para garantizar que el desarrollo y el uso de tecnologías de IA de alto riesgo respeten objetivos de interés general, como la protección de la salud, la seguridad y los derechos de los consumidores.

Asimismo, clasifica los sistemas de IA según el nivel de riesgo que representan: cuanto mayor sea el riesgo, más estricta será la regulación. Incluso los sistemas de riesgo mínimo deben ser evaluados individualmente. La prioridad del Parlamento Europeo es garantizar que los sistemas de IA en la Unión sean seguros, transparentes, trazables, no discriminatorios y respetuosos con el medio ambiente. Para ello, se exige que su supervisión

recaiga en personas y no en máquinas, reduciendo así el riesgo de consecuencias perjudiciales.

El Real Decreto-Ley 6/2023, de 19 de diciembre, por el que se aprueban medidas urgentes para la ejecución del Plan de Recuperación, Transformación y Resiliencia en materia de servicio público de justicia, función pública, régimen local y mecenazgo. En esta Ley, se habla del expediente judicial electrónico y regula la automatización de procesos o los llamados procesos inteligentes. En este sentido, se diferencian las actuaciones automatizadas, las actuaciones proactivas y las actuaciones asistidas.

En el art. 56, se denomina actuación automatizada a "[...] la actuación procesal producida por un sistema de información adecuadamente programado sin necesidad de intervención humana en cada caso singular".

En ese mismo precepto, se nos dice que actuaciones proactivas son "[...] las actuaciones automatizadas, auto-iniciadas por los sistemas de información sin intervención humana, que aprovechan la información incorporada en un expediente o procedimiento de una Administración pública con un fin determinado, para generar avisos o efectos directos a otros fines distintos, en el mismo o en otros expedientes, de la misma o de otra Administración pública, en todo caso conformes con la ley".

Por último, las actuaciones asistidas son aquellas para las que el sistema de información de la Administración de Justicia genera un borrador total o parcial de documento complejo basado en datos, que puede ser producido por algoritmos, y puede constituir fundamento o apoyo de una resolución judicial o procesal –art. 57–.

5 El desarrollo de sistemas predictivos en la prevención e investigación penal: ecosistema jurídico-tecnológico

Examinemos la evolución de los sistemas predictivos en la prevención y la investigación penal y su estado actual. En este contexto, es fundamental abordar la creación gradual de un

ecosistema jurídico-tecnológico. Desde la introducción de los primeros sistemas expertos anglosajones, conocidos como *Expert Systems* (Kalinowsky, 1973), hasta las tecnologías actuales aplicadas al proceso penal y en la prevención y la investigación, ha habido una transformación significativa.

Los *Expert Systems* son asistentes inteligentes que, utilizando modelos de computación lógica, permiten realizar un tipo de razonamiento jurídico. Inicialmente, eran herramientas simples y rudimentarias, ya que no ofrecían respuestas argumentadas a las preguntas planteadas. Con el tiempo, fueron mejorándose, incorporando capacidades analíticas e interpretativas y entrenándose con grandes cantidades de datos a través de técnicas de *deep learning* (Barona Vilar, 2019).

En este contexto, también es relevante mencionar la jurimetría, que emplea métodos cuantitativos y estadísticos para analizar el Derecho. Estos métodos permiten medir ciertos resultados y, a partir de ellos, prever nuevos escenarios, todo basado en datos judiciales. Fue en 1950 cuando Wiener, creador de la cibernética, junto con Loevinge, aplicaron la jurimetría al ámbito jurídico (Barona Vilar, 2019).

¿Qué avances aporta la jurimetría al Derecho, y en particular al sistema de justicia penal? Al analizar datos, la jurimetría permite predecir resultados. Esto mejora la gestión de los tiempos, reduce trámites innecesarios y, por lo tanto, los costes, al mismo tiempo que facilita el diseño de estrategias más eficaces, ya que las decisiones pueden tomarse basándose en el análisis de hechos pasados.

No solo han surgido sistemas tecnológicos que optimizan la toma de decisiones, sino que, en los últimos años, especialmente desde la llegada de la pandemia, hemos observado una transformación significativa en los procedimientos judiciales. La necesidad de adaptarse a la tecnología debido al confinamiento y la imposibilidad de desplazarse físicamente han provocado un cambio en la forma en que se lleva a cabo la justicia, pasando de un modelo presencial a uno más digital.

Aparece lo que se denomina *E-Justice*: la creación de sistemas que permiten agendas electrónicas conjuntas entre las Fuerzas

y Cuerpos de Seguridad del Estado y los juzgados[15]; la instauración del expediente judicial electrónico[16] y la obligatoriedad de la utilización de los medios telemáticos en la relación de los profesionales con la Administración de Justicia; la creación de la "Carpeta Justicia", que supone la implantación de un sistema de acceso único y personalizado de todo ciudadano a sus expedientes judiciales, la implementación de forma obligatoria de las vistas virtuales y la automatización del proceso judicial son señas identitarias de este concepto[17], obligando eso sí a la creación de nuevas estructuras, al reforzamiento y la modernización de la planta judicial y a la necesaria formación de los distintos operadores jurídicos.

Los jueces *robot* son ya una realidad en distintos lugares del mundo. Son varios los ejemplos que pueden ponerse: Estonia,

15. El procedimiento para el enjuiciamiento rápido de determinados delitos, previsto en el art. 795 de la LECr, y los juicios por delitos leves del art. 962, que surgieron por la Ley 38/2002, de 24 de octubre, de reforma parcial de la Ley de Enjuiciamiento Criminal, sobre procedimiento para el enjuiciamiento rápido e inmediato de determinados delitos y faltas, y de modificación del procedimiento abreviado, supusieron la creación de un sistema de "agenda común", esto es, compartida entre las Fuerzas y Cuerpos de Seguridad del Estado y los juzgados de instrucción y entre estos y los juzgados de lo penal, que permitieron agilizar las citaciones para la sustanciación de la fase de investigación y celebración de las vistas orales.

16. El expediente judicial electrónico surgió en el año 2011, a través de la Ley 18/2011, de 5 de julio, reguladora del uso de las tecnologías de la información y la comunicación en la Administración de Justicia. El Real Decreto-Ley 6/2023, 19 de diciembre, por el que se aprueban medidas urgentes para la ejecución del Plan de Recuperación, Transformación y Resiliencia en materia de servicio público de justicia, función pública, régimen local y mecenazgo, tal y como dice la Exposición de Motivos: "[…] persigue, en primer lugar, la adaptación de la realidad judicial española del siglo XXI al marco tecnológico contemporáneo, favoreciéndose una relación digital entre la ciudadanía y los órganos jurisdiccionales y aprovechando las ventajas del 'hecho tecnológico' también para fortalecer nuestro Estado social y democrático de Derecho mediante la disposición de medidas orientadas a la transparencia, la eficiencia y la rendición de cuentas de los poderes públicos". Podemos decir que esta norma supone la implementación del expediente judicial electrónico 2.0.

17. Todo ello a través del Real Decreto-Ley 6/2023, 19 de diciembre, por el que se aprueban medidas urgentes para la ejecución del Plan de Recuperación, Transformación y Resiliencia en materia de servicio público de justicia, función pública, régimen local y mecenazgo.

Argentina, Colombia[18]. Por otra parte, la automatización de la justicia es una realidad ya en España y los jueces ya están utilizando la IA como auxilio en su actividad jurisdiccional. De hecho, ha sido reciente cuando un juez de los Países Bajos, concretamente, el titular del Tribunal de primera instancia de *Gelderland*, resolviendo un proceso civil, concretamente un conflicto entre vecinos sobre la instalación de paneles solares y una estructura de techo adyacente, ha recogido en su sentencia cuál es, según el *ChatGPT*, la esperanza de vida media restante de unos paneles solares y el precio promedio de la electricidad en ese momento, datos esenciales para determinar de forma precisa la cuantía de la compensación económica a la que debía ser condenada la parte demandada[19]. Observemos, pues, cómo ha cambiado la forma de fundamentación de las resoluciones judiciales.

Hoy, en día, los profesionales del Derecho cuentan con herramientas de IA generativa que les permiten realizar tareas legales de manera más rápida, reduciendo costes y errores. Estas herramientas incluyen asistentes para redactar y analizar documentos legales, preparar vistas y pruebas, organizar auditorías y gestionar despachos, lo que ya es una realidad en la práctica actual. Asimismo, las Fuerzas y Cuerpos de Seguridad del Estado han transformado su forma de actuar para prevenir e investigar delitos; ya no se basa en la intuición, sino en el uso de herramientas

18. En Estonia, un juez *robot* decide asuntos de no más de 7.000 €. En Argentina, está *Prometea*, que se encarga de la resolución de infracciones menores en materia de tráfico, por ejemplo. En Colombia, está *Pretoria*, que resuelve también casos urgentes en la Corte Constitucional, aunque con supervisión humana.

19. El juez explica en su sentencia que tales datos los obtuvo consultando el *ChatGPT*, pero el problema es que justificó la condena y la cuantía de la indemnización sobre unos datos no proporcionados por un perito experto en la materia, sino sobre elementos calculados por un *robot*, desconociendo cuáles son los algoritmos utilizados para ello. Las preguntas que cabe realizarse son dos: primera ¿puede un sistema robótico no especializado dar datos fiables, sustituyendo un dictamen pericial y justificar estos la decisión de un juez? La cuestión no es nueva: estamos en el llamado "conocimiento privado del juez". El problema es la indefensión que esto puede llegar a generar. Si alguna de las partes quisiera recurrir, no podría atacar, por desconocer los datos con los que ha sido entrenada la IA, el algoritmo con arreglo al cual ha calculado esas cifras y, por ende, no podría discutir la valoración de la prueba y no tendría cómo impugnar la argumentación de la resolución judicial.

predictivas de prevención e investigación, de las que hablaremos a continuación.

6 La prevención, la investigación y la justicia penal predictiva: la aplicación de la IA en estos ámbitos

Es posible poner muchos ejemplos de cómo se trabaja en el ámbito de la vigilancia y la prevención delictual con herramientas de IA predictivas. Ha surgido así la denominada *predictive policing* o "justicia predictiva policial", o vigilancia predictiva (Perry, McInnis, Price, Smith y Hollywood, 2013) y, por ende, la criminología ambiental o criminometría: métodos cuantitativos de análisis en el ámbito policial, utilizados para identificar objetivos, planificar la actividad policial, prevenir los delitos y resolver casos del pasado, a través de sistemas que permiten predecir estadísticamente lo que va a suceder (Barona Vilar, 2019). En palabras de la Organización para la Seguridad y la Cooperación en Europa, este fenómeno consiste en "la recopilación y evaluación sistemática de datos e información, a través de un proceso analítico definido, que los convierte en productos analíticos estratégicos y operativos, que sirven de base para un proceso decisorio mejorado, fundamentado y documentado"[20].

Se utilizan foros, webs, redes sociales y aplicaciones móviles para identificar áreas de alto riesgo y crear, así, los llamados "puntos calientes" o *hot spots*. La ventaja de detectar estos puntos es la posibilidad de elaborar mapas digitales de delitos, lo que permite conocer las zonas más propensas para la comisión de ciertos crímenes. Esto facilita la planificación de medidas preventivas, el refuerzo de la seguridad y la distribución eficiente de recursos policiales y materiales. Los mapas de riesgos se han vuelto esenciales para la prevención delictiva.

Estos sistemas predictivos comenzaron a utilizarse en EE. UU. En, Chicago, en 1920, se creó un *software* llamado BIG DATA.

20. Véase la Guía de la OSCE sobre actividad policial basada en la inteligencia, 2017, p. 6. https://www.osce.org/files/f/documents/6/4/455536.pd (Consultado 09/09/2024).

Fueron los orígenes de la aplicación de la IA a la predicción delictual.

En Europa, las primeras herramientas de análisis predictivo se utilizaron en Francia en 1994 con *Anacrim*, sistema que fue reemplazado en 2005 por *i2 Analyst Notebook (i2AN)*, creando una base de datos estatal para compartir información entre distintos órganos. Estas herramientas permiten establecer conexiones entre personas y delitos, algo que el ser humano no puede hacer. En Francia, existen sistemas como *Chardon* y *Salvac* para identificar delitos violentos o sexuales cometidos por la misma persona.

En Italia, en 2007, se implementó *KeyCrime* en Milán para predecir crímenes en serie. En el Reino Unido, en 2013, se usó *PredPol* en *Kent* para la prevención delictiva. Bélgica y los Países Bajos adoptaron el sistema *Crimen Anticipation System* (CAS), mientras que en Alemania, en 2015, se implementó *Skala*, que analiza factores socioeconómicos y redes de comunicación para predecir la probabilidad de fuga de un delincuente y detectar zonas con alta delincuencia.

En España, la Policía Municipal de Madrid y la Policía Nacional utilizan el Sistema de Información Geográfica (*SIG*), que, desde 2015, ayuda en la geoprevención y permite crear estrategias preventivas para reducir la delincuencia, integrando datos sobre la relación entre los agentes del crimen y el territorio.

Hablemos también de *CATT*, en España, que es un sistema que busca identificar por el análisis del lenguaje el discurso utilizado por los abusadores y *SWEETIE*, en Australia, que ha permitido, a través de una niña virtual que aparece en *chats* y *webs* de citas, detectar pedófilos. *VERIPOL* es un sistema de IA utilizado por las Fuerzas y Cuerpos de Seguridad del Estado español y creado en colaboración con la Universidad Complutense de Madrid, que, a través de métodos de procesamiento del lenguaje natural y aprendizaje automático, posibilita calcular la probabilidad de que una declaración o denuncia sea falsa[21].

21. Véase https://www.ucm.es/otri/veripol-inteligencia-artificial-a-la-caza-de-denuncias-falsas (Consultado 29/10/2024).

En el ámbito de la investigación criminal, se deben citar las herramientas de investigación de los Cuerpos y Fuerzas de Seguridad del Estado. Así, *VALCRI*, *Visual Analytics for sense making in Criminal Intelligence Analysis*. Esta IA sirve a los investigadores en sus labores para relacionar evidencias habidas en la escena del crimen con datos obrantes en las bases de datos de la Policía y basados en técnicas biométricas y reconocimiento facial[22].

En relación con los modelos biométricos, existen sistemas de IA que permiten la identificación de huellas dactilares, geometría de la mano, identificación del iris, imagen fácil, otograma (reconocimiento de la oreja), etc. Tienen por finalidad reconocer y autenticar a las personas que reúnen una serie de características fisiológicas o morfológicas determinadas (Boulgouris, 2010). Estos sistemas son aplicables en materia de investigación penal predictiva.

En primer lugar, es preciso tener en cuenta el concepto que el Reglamento de la UE sobre IA nos proporciona tanto sobre datos biométricos como sobre identificación biométrica y verificación biométrica, según el art. 3.

Los datos biométricos son "los datos personales obtenidos a partir de un tratamiento técnico específico, relativos a las características físicas, fisiológicas o conductuales de una persona física, como imágenes faciales o datos dactiloscópicos".

La identificación biométrica es "el reconocimiento automatizado de características humanas de tipo físico, fisiológico, conductual o psicológico para determinar la identidad de una persona física comparando sus datos biométricos con los datos biométricos de personas almacenados en una base de datos". Esta puede ser remota, que es la que permite identificar a las personas físicas sin su participación activa y generalmente a distancia, comparando sus datos biométricos con los que figuran en una base de datos de referencia y esta a su vez, en tiempo real

22. Véase Jiménez Conde, F. y Bellido Penadés, R. (coords.) (2019). *Justicia: garantías vs. Eficacia* (pp. 665-674). Valencia: Tirant lo Blanch.

o en diferido. La primera es aquella en la que la recopilación de los datos y la comparación y la identificación tienen lugar sin una demora significativa; pudiendo ser identificación instantánea y no instantánea; esta última cuando se produce una demora mínima limitada. En diferido, es la identificación que no se produce en tiempo real.

La verificación biométrica será automatizada y uno-a-uno, incluida la autenticación, de la identidad de las personas físicas mediante la comparación de sus datos biométricos con los datos biométricos facilitados previamente.

En este sentido, mencionamos que la Policía Nacional lleva ya unos cuantos meses utilizando un sistema de reconocimiento biométrico, llamado *ABIS*, que permite realizar un reconocimiento facial de una persona si sobre ella existen registros. En total, hay trece estaciones operativas de *ABIS* repartidas por el país: dos, en Madrid y una, en Barcelona, Granada, Málaga, Sevilla, Valencia, Valladolid, Las Palmas, Zaragoza y Bilbao. Pronto se unirá otra en Pamplona. Asimismo, la Guardia Civil cuenta con dos estaciones de reconocimiento facial en Madrid y los *Mossos d'Esquadra* también están implementando un sistema parecido[23]. Estos métodos suponen una revolución en los procedimientos aplicados por la Policía en la investigación de los delitos, pues la única posible identificación de una persona presuntamente responsable de un delito hasta el momento era la que se producía a través de una prueba pericial consistente en el análisis de la huella dactilar o de *ADN*.

El sistema IA funciona de la siguiente manera: en una primera fase, se extrae el rostro de la imagen mediante una tecnología llamada visión computacional. Tras ello, se aplica un algoritmo a ese rostro, obteniendo un patrón que lo represente y lo diferencie de los demás. El algoritmo posibilita buscar ese patrón en extensos bancos de imágenes y ofrecer los resultados que más se parezcan. Estos sistemas de reconocimiento facial pueden ser aplicados donde antes se hacían retratos *robot*.

23. https://elpais.com/tecnologia/2024-05-28/la-policia-espanola-ya-usa-en-sus-investigaciones-un-sistema-automatico-de-reconocimiento-facial.html (Consultado 29/10/2024).

En la UE, el uso de sistemas de identificación biométrica remota en tiempo real en espacios públicos está prohibido, salvo cuando sea estrictamente necesario para objetivos específicos, como la búsqueda de víctimas de secuestros o personas desaparecidas, o para prevenir amenazas graves e inminentes a la vida o la seguridad. En estos casos, solo se permitirá para confirmar la identidad de la persona objetivo, teniendo en cuenta la gravedad de la situación y el impacto en los derechos y libertades de las personas involucradas. Además, se deben cumplir garantías y condiciones proporcionales según el Derecho nacional, como limitaciones temporales, geográficas y personales. Estos sistemas deben ser autorizados y registrados en la base de datos de la UE, excepto en casos de urgencia justificada. Los Estados miembros pueden autorizar, total o parcialmente, el uso de estos sistemas bajo las condiciones mencionadas. También están prohibidos los sistemas de IA que creen bases de datos de reconocimiento facial mediante la recolección no selectiva de imágenes de internet o circuito cerrado de televisión –CCTV–.

¿Cuál es la razón de tales restricciones previstas en el Reglamento de IA? Se ha demostrado que los sistemas de reconocimiento biométrico se han utilizado con fines discriminatorios y racistas[24]. Si a esto se le une la forma que utilizan las herramientas de IA para disponer de una base de datos de millones de rostros, el problema se agrava aún más, pues ciertos *softwares* aprovechan todas las imágenes publicadas en perfiles y páginas *web* públicas para hacerse con ellas y poder utilizarlas en los posteriores reconocimientos. La vulneración de derechos fundamentales, tales como la privacidad y el derecho a la propia imagen, es flagrante, pues todo ello se hace sin consentimiento de los titulares[25].

Hablando ya del ámbito judicial, que no policial, y en el de la llamada justicia penal predictiva, destacan sistemas como el

24. Véase el Informe emitido por Amnistía Internacional en relación a los sistemas de reconocimiento facial implementados en la ciudad de Nueva York. https://www.amnesty.org/es/latest/news/2022/02/usa-facial-recognition-technology-reinforcing-racist-stop-and-frisk-policing-in-new-york-new-research/ (Consultado 09/10/2024).

25. *Clearview* es una herramienta de IA de reconocimiento facial https://www.clearview.ai/ (Consultado 09/10/2024). En su propia página *web*, se anuncian como colaboradores de las Fuerzas y Cuerpos de Seguridad del Estado para garantizar la seguridad ciudadana. Ellos mismos indican que contribuyen a la "caza" de depredadores sexuales.

ya mencionado de *COMPAS* (EE. UU.) y *HART* (Reino Unido). El primero es un sistema que utiliza, como se indicó, técnicas de aprendizaje automático, dentro de un sistema de aprendizaje supervisado, concretamente un sistema de clasificación, para predecir la probabilidad de que un individuo reincida o cometa un delito en el futuro, lo que se traduce en una tarea de clasificación binaria –reincidencia o no reincidencia–. El algoritmo utilizar técnicas de aprendizaje supervisado para aprender patrones en los datos de entrenamiento y luego aplica esos patrones para hacer predicciones sobre nuevos individuos. *HART* también es capaz de pronosticar si los sospechosos tienen un bajo, moderado o alto riesgo de cometer más delitos en un periodo de dos años. Ambos sistemas se aplican en la adopción de medidas cautelares como la privación de libertad o sistemas de rehabilitación. Por ejemplo, *HART* utiliza datos de 34 categorías diferentes (edad, sexo, domicilio, antecedentes penales, profesión, estado civil, etc.)[26].

El sistema de IA llamado *LSI-R*, *Level of Service Inventory-Revised*, se utiliza en los permisos de salida y la libertad condicional de los procesados, basándose en la ponderación de criterios tales como antecedentes penales, lugar de residencia, educación, empleo, ocio, familia, problemas de alcohol o drogas, actitudes emocionales y personales.

RIS CANVI es un sistema utilizado en Cataluña por los jueces de instituciones penitenciarias y en virtud del cual pueden otorgar permisos de salida de los presos.

Otra herramienta similar a las anteriores es *VIOGÉN*, que se utiliza en España para analizar los riesgos de reincidencia en el ámbito de la violencia de género.

En el ámbito de la UE, y según el Reglamento de IA, están prohibidos los sistemas de IA para realizar evaluaciones de riesgos de personas físicas con el fin de valorar o predecir el riesgo de que una persona física cometa un delito, cuando se basan de forma exclusiva en la elaboración de su perfil o en la evaluación

26. Véase https://www.durham.police.uk/Information-and-advice/Pages/Checlpoint.aspx (Consultado 09/10/2024).

de los rasgos y características de su personalidad, aunque dicha prohibición no aplica cuando sirvan de apoyo a la valoración humana de la implicación de una persona en una actividad delictiva que ya se base en hechos objetivos y verificables directamente relacionados con una actividad delictiva –art.3.1.d)–.

Pero no solo existen herramientas de IA aplicables a medidas cautelares, también se han desarrollado ya métodos cuantitativos que permiten una algoritmización de la prueba[27]. Posibilitan un análisis de riesgos, por ejemplo, sobre la fiabilidad de un testigo. Recordemos que la prueba testifical se valora según criterios lógicos y de razón del juez, debiendo este motivar en la sentencia la fijación como ciertos de los hechos en virtud de la prueba testifical realizada. En este sentido, cabe mencionar *ADVOCATE*.

En otro orden de cosas, también se han creado sistemas de IA que ayudan a los jueces a dictar sentencias, es el caso de *ASSYST*, en Canadá (Simón y Gaes, 1989) o *LIST*, en Columbia (Schild, 1998). Todos ellos son herramientas predictivas aplicadas en la función jurisdiccional.

¿Qué se entiende, entonces por justicia penal predictiva? Son términos muy amplios. En definitiva, el conjunto de herramientas que buscan la eficiencia procesal, que supone una mejora en la calidad de la toma de decisiones y una aminoración del trabajo realizado por los jueces (Armenta Deu, 2021).

7 Fortalezas, debilidades y oportunidades

El uso de sistemas de inteligencia artificial predictiva facilita la creación de un sistema preventivo de seguridad y lucha contra el delito, basado en un análisis de riesgos mucho más preciso

27. Además de las herramientas de IA predictivas, también existen aquellas que podemos incluir dentro de la IA generativa. Ya hay las que permiten almacenar datos y configuran un documento específico, que puede utilizarse como prueba documental. O aquellas otras que crean informes-auditorías que pueden también usarse con valor probatorio equivalente al dictamen pericial (Barona Vilar, 2019).

que el realizado a través de cálculos humanos. Los errores humanos en la valoración de datos se reducen al utilizar algoritmos automáticos que permiten prever lo que podría ocurrir, ajustando los recursos necesarios en consecuencia.

En términos de seguridad y prevención delictiva, estos sistemas optimizan tanto los recursos materiales como humanos, asignando recursos de acuerdo a las necesidades reales. Además, facilitan la vigilancia dinámica, mejorando su calidad visual y acústica, y permiten una replanificación de los recursos cuando los inicialmente previstos no sean suficientes o adecuados. Todo ello contribuye a un ahorro de costes y a una mejora en la seguridad ciudadana, optimizando los recursos aplicados en el proceso penal, que debe ser eficiente y debe ser considerado como servicio público.

Una planificación eficaz en la prevención delictiva impacta directamente en la estructuración del Derecho Penal como *ultima ratio*, respaldando la idea de que "más vale prevenir que curar". Con ello, se prioriza un Derecho Penal *ex ante*, que otorga mayor protagonismo a las Fuerzas y Cuerpos de Seguridad del Estado en la prevención, en lugar de la represión, tras la comisión de un delito, tras la sustanciación del proceso penal. Así, se actúa ante los riesgos y las amenazas, en vez de centrarse en la sanción penal. Reforzar la fase de prevención preprocesal es una característica de los sistemas avanzados.

El cálculo de riesgos realizado por la IA, en cuanto a la probabilidad de reincidencia o la evaluación de la credibilidad de un testigo, permite ahorrar tiempo a los jueces al basarse en criterios matemáticos para sus decisiones. La automatización de estos procesos aumenta la eficiencia, dejando que los jueces realicen tareas que requieren su intervención directa. Además, el cálculo matemático de los riesgos proporciona objetividad, promoviendo la imparcialidad en decisiones relacionadas con los derechos fundamentales.

En cuanto a las pruebas penales, la IA predictiva mejora el sistema al complementar las pruebas periciales, permitiendo probar hechos que anteriormente no era posible. Esto contribuye a un proceso penal más moderno y acorde con los tiempos,

permitiendo la utilización de nuevas figuras como la vigilancia dinámica o el agente encubierto informático.

Por último, debe plantearse también la ventaja que tiene el posible uso de la IA predictiva por parte de los abogados. Ahorro de tiempo y eficiencia son dos de las fortalezas que para un letrado puede representar el uso de esta tecnología. Si un algoritmo le calcula el riesgo de reincidencia de un cliente, le será mucho más fácil plantear la estrategia de defensa frente a estos cálculos.

Sin embargo, a pesar de sus fortalezas, el uso de la IA predictiva presenta importantes debilidades. La IA no es infalible y, aunque en principio es neutral, puede sufrir sesgos y ser manipulada con fines malintencionados. Los algoritmos, lejos de ser imparciales, pueden generar discriminación. Si las Fuerzas y Cuerpos de Seguridad del Estado, si los jueces utilizan estas herramientas debemos estar seguros de que no se basan en cálculos matemáticos discriminatorios. Por ello, debe regularse que los algoritmos aplicados en el ámbito del Derecho y, concretamente, en la predicción penal debe estar controlados por una autoridad pública.

Debemos advertir sobre la utilización en remoto de los reconocimientos biométricos o faciales. El fin no justifica los medios. La prevención y la investigación criminal no deben suponer quebrantar derechos fundamentales, tales como la intimidad o el derecho a la propia imagen de las personas. Atención debe prestarse a las fuentes de las que se nutren las bases de datos que después se toman como muestra para realizar las identificaciones.

Además, estos sistemas predictivos pueden vulnerar derechos fundamentales, como el derecho de defensa, la presunción de inocencia, el principio *in dubio pro reo* y el principio de contradicción. Si la IA se utiliza como única prueba, la sentencia de condena podría carecer de base suficiente para invalidar la presunción de inocencia, ya que los algoritmos pueden presuponer culpabilidad. La opacidad de los algoritmos, o las llamadas "cajas negras", dificulta la comprensión de los resultados obtenidos.

La falta de motivación en las resoluciones judiciales basadas únicamente en herramientas predictivas también plantea un problema, ya que la motivación es esencial para garantizar el derecho a la tutela judicial efectiva. La ausencia de una justificación suficiente puede vulnerar este derecho, como ha señalado el Tribunal Constitucional.

Una debilidad crítica es la responsabilidad por los daños causados por predicciones erróneas. Debe aclararse quién es responsable: ¿el que eligió el algoritmo?, ¿el que seleccionó los datos con los que la IA fue entrenada?, ¿quién no revisó los resultados?, ¿quién los interpretó o los aplicó?

En este contexto, resulta relevante la sentencia de la Audiencia Nacional de 30 de septiembre de 2020, que establece que la información obtenida en la fase preprocesal de una investigación policial puede orientar al juez en la adopción de medidas cautelares, pero no es decisiva. Se trata de un asesoramiento especializado que ayuda en la valoración del "riesgo objetivo para la víctima", entre otros instrumentos incluidos en la LECrim para tomar decisiones durante la instrucción judicial.

Otra debilidad significativa es la responsabilidad por errores predictivos. Se debe establecer quién asume la responsabilidad en caso de que los resultados erróneos de la IA causen daños. En este contexto, la jurisprudencia sugiere que la IA debe considerarse como una herramienta de asesoramiento, no como la base decisiva de una sentencia, especialmente en la adopción de medidas cautelares.

El hecho de que los jueces puedan tener dudas no implica que su valoración deba ser sustituida por los resultados arrojados por un algoritmo o una máquina. En fase cautelar, encontramos el problema de la presunción de inocencia. La doctrina está dividida. Mientras que unos autores manifiestan sus reticencias (Llorente Sánchez-Arjona, 2022), otros indican que la ayuda proporcionada por la IA no debe de ser desechada: las medidas cautelares son necesarias y siempre se adoptan valorando determinados riesgos –fuga, reiteración delictiva o de comisión de delito frente a bienes jurídicos de la víctima–, sin que haya motivo para restringir sin más los elementos cognitivos que

utiliza el juez a la hora de dictar un auto de prisión provisional (Hoyos Sancho, 2020).

La misma LECrim proporciona los elementos que han de tenerse en cuenta a la hora de valorar los requisitos de la prisión provisional. Por ejemplo, para valorar la existencia del riesgo de fuga –art 503.1.3.º LECrim y, por todas, sentencia del Tribunal Constitucional 128/1995, de 26 de julio–, se debe atender de forma conjunta a los siguientes elementos: la naturaleza del hecho; la gravedad de la pena que pudiera imponerse al investigado o encausado; la situación familiar, laboral y económica de este, y la inminencia de la celebración del juicio oral. La valoración del juez se realiza antes de adoptar una decisión sobre una medida cautelar, no pudiendo obviar que el juez es profesional y cuenta con experiencia, siendo imparcial e independiente. Argumentar que el juez puede tener sesgos a la hora de tomar decisiones es tanto como poner en duda su imparcialidad. Los sesgos que puede tener un juez no serán evitados por la IA. Simón Castellano (2021) pone de manifiesto que estos sesgos se podrían replicar en el caso de una máquina, pudiéndose producir incluso automatismos que llevasen a perpetuar los producidos por un juez humano o incluso agravarlos, al programar la máquina de acuerdo a los patrones de decisión humanos, sin volver a analizarse el nivel de riesgo y la proporcionalidad de la medida, de forma individualizada y según las circunstancias de cada caso concreto.

La aplicación de la IA en fase de ejecución plantea menos problemas, al existir una sentencia firme de condena.

En relación con todo lo expuesto debemos citar el informe *Artificial Intelligence and Fundamental Rights European Union Agency for Fundamental Rights* (2020), el cual se refiere expresamente a la justificación adecuada de los criterios y procesos mediante los cuales se adoptan decisiones basadas en algoritmos.

Decíamos que la utilización de la IA para los abogados también representa una ventaja, ¿ningún riesgo? Pues sí y es que el uso de la IA plantea importantes cuestiones relacionadas con el derecho de defensa y el principio de igualdad de armas. Si una parte tiene acceso a tecnología de IA avanzada mientras que la otra no, se puede producir un "desequilibrio cognoscitivo" que

afecta la paridad entre las partes. Este desequilibrio se manifiesta principalmente en dos aspectos: acceso a la tecnología, pues el Ministerio Fiscal generalmente tendrá acceso a tecnología más moderna y recursos económicos que no están al alcance del letrado del acusado. En segundo lugar, la comprensión de la evidencia, pues, si resulta imposible acceder al "código fuente" del algoritmo de IA utilizado, sería casi imposible para la defensa cuestionar o impugnar los resultados proporcionados por el sistema que podrían usarse como prueba. Para garantizar el principio de igualdad de armas, se precisaría que el Estado garantice el acceso a esta tecnología a la parte que no pueda costearla por sí misma. Ha de decirse que valorar la prueba obtenida basada en algoritmos proporcionados por la IA y su posible refutación, cuando existe asimetría tecnológica, supone una vulneración del derecho de defensa, del principio de contradicción y de igualdad. Se precisa asegurar que todas las partes intervinientes en el proceso puedan comprender, analizar y cuestionar las evidencias generadas por sistemas de IA para mantener un proceso justo y equitativo.

7.1 ¿Cuáles son las oportunidades?

Una de las principales oportunidades es la capacidad de predecir, basándose en datos objetivos y matemáticos, las tendencias delictivas, lo que permite identificar patrones y anticipar comportamientos delictivos. Esta capacidad de anticipación es valiosa en cualquier área, pero resulta aún más relevante en el ámbito de la delincuencia.

Al evaluar los riesgos, se mejora la toma de decisiones, facilitando la identificación de áreas que requieren intervención. Esto es especialmente útil en la planificación de sistemas de rehabilitación y para la personalización de la supervisión. Además, ayuda a identificar qué factores deben ser protegidos, permitiendo definir áreas prioritarias de atención, tanto a nivel policial como judicial.

El uso de sistemas algorítmicos representa una oportunidad para fortalecer la imparcialidad de los jueces, ya que sus decisiones se basarán en datos precisos y objetivos. Este enfoque

contribuye a corregir posibles sesgos humanos y reafirma la equidad en la toma de decisiones judiciales.

Actualmente, en España no existen unas normas éticas de uso de la IA ni predictiva ni generativa. Sin embargo, sí están los principios procesales básicos, las normas procesales sobre la prueba y el código deontológico de las distintas profesiones, tal y como es el Código deontológico de la Abogacía española, aprobado por el Pleno del Consejo General de la Abogacía Española el 6 de marzo de 2019 o el Código deontológico de los procuradores de los tribunales, aprobado por el Consejo General de Procuradores de España.

En España, recientemente, el 21 de junio de este año, se ha aprobado la política de uso de la IA en la Administración de Justicia en la Secretaría General de Comité Técnico Estatal de la Administración Judicial Electrónica[28], aunque todavía debe ser ratificada por el Consejo General del Poder Judicial –en adelante, CGPJ– y por la Fiscalía General del Estado –en adelante, FGE–, por las comunidades autónomas con competencia en Justicia, Ministerio de la Presidencia y Ministerio de Justicia. Este documento afecta a 1.400 jueces y magistrados que forman parte del Poder Judicial.

Debe tenerse en cuenta que las Fuerzas y Cuerpos de Seguridad del Estado, así como los órganos jurisdiccionales y los jueces y magistrados, garantes de la tutela judicial efectiva, deben cumplir el Reglamento de IA de la UE, la norma de protección de datos personales tanto de la UE como española. Por eso, deben adoptarse criterios mínimos, tanto para los que impulsan los proyectos como para los propios usuarios.

Los destinatarios de esta guía son todos los trabajadores de la Administración de Justicia, el personal al servicio de proveedores de herramientas de inteligencia artificial, así como cualquier otro actor institucional, público o privado, que tenga acceso a la información que obre en poder de las administraciones con competencias y CGPJ, o se encuentre alojada en los sistemas destinados a la Administración de Justicia.

28. https://www.administraciondejusticia.gob.es/cteaje/normativa-complementaria (Consultado: (Consultado 24/09/2024).

En esta guía, se trazan unas líneas rojas que no pueden traspasarse en este ámbito, como son los derechos fundamentales, el principio de no discriminación, de calidad y seguridad, respeto a la transparencia, imparcialidad y lealtad, control del usuario, equidad y acceso universal, prevención de sesgos y discriminación, protección de la privacidad y datos personales, innovación responsable y evaluación continua, formación y capacitación y cogobernanza.

Debe quedar claro que la IA nunca puede reemplazar a las decisiones humanas. Es una herramienta de ayuda, pero la responsabilidad final es siempre, en este caso, de los jueces, magistrados y letrados de la Administración de Justicia. Sus decisiones tienen que ser independientes.

El desarrollo y la aplicación de la IA no pueden ser discriminatorios, dado que su uso supone el conocimiento y la utilización de datos sensibles.

Los modelos y algoritmos creados se almacenarán y ejecutarán en entornos seguros, para garantizar la integridad del sistema y su intangibilidad. Cuando se utilicen fuentes certificadas, no podrán ser modificadas hasta que hayan sido utilizadas en el mecanismo de aprendizaje, siendo necesario que todo el proceso sea rastreable para que no se produzca ninguna alteración.

Es preciso lograr un equilibrio entre la propiedad intelectual y los principios de transparencia, imparcialidad, equidad y lealtad, para que no se produzcan sesgos, priorizando el interés de la justicia. Resulta imprescindible que los algoritmos sean transparentes y las operaciones automatizadas sean comprensibles.

Se recomienda ofrecer publicidad y transparencia a través de páginas *web* oficiales. Por ejemplo, se aconseja publicar los registros FAT *(Fairness, Accuracy and Transparency)* o similares, acerca de los datos usados, miembros de los equipos de IA, servicios, algoritmos, posibles sesgos y aplicaciones que hacen uso de técnicas de inteligencia artificial.

Debe garantizarse que, aunque haya una actuación de la IA, esta pueda ser revisable previamente, informando a los ciudadanos, de

una forma clara y comprensible, sobre si los resultados ofrecidos por la IA son vinculantes, su uso en el procedimiento judicial y su derecho a objetar, pudiendo ser oído directamente por un órgano judicial.

Debe garantizarse un acceso equitativo a los sistemas judiciales, independientemente de la ubicación, el estatus socioeconómico o cualquier otra característica demográfica. Por ello, deben eliminarse las barreras que lo impidan.

Deben prevenirse y corregirse los posibles sesgos. Para ello, los algoritmos deben ser revisados de forma periódica.

Debe potenciarse la innovación responsable en el desarrollo y la implementación de la IA en la Administración de Justicia. Esto supondrá evaluaciones periódicas del impacto de la tecnología en el sistema judicial, debiendo realizarse ajustes y mejoras, para garantizar su efectividad y equidad.

En este principio está uno de los grandes escollos de la aplicación de la IA en el ámbito de la Administración de Justicia, dado que debe proporcionarse formación y capacitación adecuada a los profesionales del Derecho y a todos los actores que usan esta tecnología, insistiendo en su componente ético. ¿Por qué es una dificultad? Porque los profesionales de la justicia son reticentes a la formación sobre cuestiones tecnológicas, no por falta de interés, sino de tiempo.

El principio de cogobernanza significa colaboración, compartiendo e intercambiando conocimientos, así como los propios sistemas basados en IA entre diferentes áreas de la organización, o con el resto de las administraciones con competencias en materia de Justicia, el CGPJ y la FGE o con otras instituciones, de tal forma que se impulse el deseado desarrollo innovador en sintonía con una implementación ética de la inteligencia artificial.

Debemos alabar el hecho de la elaboración de esta guía. Sin embargo, en nuestra opinión, debería ser incorporada a preceptos normativos en la Ley Orgánica del Poder Judicial y en las diferentes leyes procesales. Cierto es que los principios enunciados en la guía se encuentran en su mayoría contemplados ya

en el espíritu de las ya existentes, incluidos los derechos fundamentales. Pero somos de la opinión de que debe preverse un sistema de límites y a su vez un régimen de responsabilidad y de sanciones en el caso de traspasarlos. Tan necesario como lo anterior, entendemos resulta imprescindible crear una cultura *behaviour*, tanto en el ámbito público como privado. Se habla ya de *behavioral compliance*, esto es, de aquella forma de pensar que entiende la ética como punto fundamental en lo que la IA se refiere.

8 Conclusiones

Los riesgos que la IA predictiva ponen encima de la mesa para los derechos fundamentales de los ciudadanos nos hacen exigir una regulación procesal española en relación a ciertas cuestiones de vital importancia. Cierto es que el Real Decreto-Ley 6/2023, 19 de diciembre, ya regula la automatización de procesos, pero resulta imperioso regular el uso de la IA en la Ley Orgánica del Poder Judicial y en las distintas leyes procesales. Algunos autores entienden que son suficientes los códigos de buenas prácticas.

1. En el ámbito policial de la prevención y la investigación predictiva, tiene que garantizarse que estos sistemas no supongan un tratamiento discriminatorio y, por tanto, que los algoritmos matemáticos puedan ser controlados por una autoridad pública, que garantice la no vulneración de derechos fundamentales. También, es aplicable esta conclusión en lo que a los jueces se refiere. Cualquier sesgo detectado debe conllevar la reeducación de los sistemas de IA que tenga por objeto su eliminación.

2. Debe prohibirse el uso indiscriminado y masivo y en remoto de los reconocimientos biométricos o faciales. Dicha prohibición debe de ser incluida en la LECr. Al mismo tiempo, las herramientas de IA que permiten estos sistemas deben estar controladas en cuanto al respeto a la privacidad de las fuentes de las que se nutren, para crear las bases de datos que después

se toman como muestra para realizar las identificaciones. Somos partidarios de prever sanciones, incluso penales, en el caso de que se vulneren derechos fundamentales cuando se realizan estos controles biométricos.

3. El art. 117 de la Constitución española –en adelante, CE– encomienda de forma exclusiva a los jueces y magistrados la función que consiste en juzgar y hacer ejecutar lo juzgado. Nada impide que los jueces se auxilien de herramientas de IA para calcular determinados datos o riesgos de los que depende su decisión. Evidentemente, y tal y como indica el Reglamento de IA de la UE, debe prohibirse categóricamente la posibilidad de que sea una máquina la que adopte por sí sola una decisión judicial. La función jurisdiccional se ejerce de forma exclusiva por los jueces y magistrados tal y como indica el art. 117.3 CE, en consecuencia, descartamos la posibilidad de que existan decisiones judiciales encomendadas a un algoritmo jurídico (Borges Blazquez, 2021 y Montesinos García, 2022). En el caso de que sean los algoritmos los que tomen decisiones en el ámbito procesal, estas decisiones deben poder ser revisadas a través de un sistema de recursos ante un juez.

4. Al mismo tiempo, debe regularse pormenorizadamente en qué actos pueden auxiliarse los jueces de la IA predictiva, por ejemplo, en la valoración de la prueba testifical, a la hora de dictar los autos relativos a las medidas cautelares, etc. Creemos necesario que se permita la utilización de estos mecanismos en la LECr, pero también se limite sus casos de uso. Al mismo tiempo, creemos que deben estandarizarse los parámetros que la IA debe tener en cuenta para realizar sus cálculos. Por ejemplo, qué parámetros deben utilizarse para calcular el riesgo de huida de una determinada persona a la que se pretende ingresar en prisión, pudiendo utilizarse los proporcionados por la jurisprudencia.

5. En todo caso, debe garantizarse que la IA sea pública y accesible y prever la obligación de que toda decisión en la que intervenga un algoritmo haga públicos el origen y los datos con los que ha sido entrenado, las reglas introducidas para calcular los datos, con la exigencia de la explicación en lenguaje claro y sencillo de aquellas, con la finalidad de que

la parte afectada pueda impugnar la decisión judicial que ha sido adoptada a partir de dichos resultados.

6. Se precisa asegurar que todas las partes intervinientes en el proceso puedan comprender, analizar y cuestionar las evidencias generadas por sistemas de IA para mantener un proceso justo y equitativo. Debe asegurarse por la legislación y, para ello, debe regularse dentro del derecho a la asistencia jurídica gratuita la posible utilización de herramientas de IA para todas las partes del proceso.

7. Debe establecerse la obligación de que todo juez que adopta una decisión basada en IA predictiva informe de ello a las partes afectadas.

8. Debe regularse quién es el responsable en caso de resultados predictivos erróneos.

9. Debe preverse un sistema de límites y a su vez un régimen de responsabilidad y de sanciones en el caso de traspasarlos.

Tan necesario como lo anterior, entendemos resulta imprescindible crear una cultura *behaviour*, tanto en el ámbito público como privado *–behavioral compliance–*.

Referencias bibliográficas

Alpaydin, E. (2014). *Introduction to Machine Learning*. Massachusetts. 3.ª edición.

Amnistía Internacional. https://www.amnesty.org/es/latest/news/2022/02/usa-facial-recognition-technology-reinforcing-racist-stop-and-frisk-policing-in-new-york-new-research/ (Consultado 9/10/2024).

Armenta Deu, T. (2021). *Derivas de la justicia* (p. 262). Madrid: Marcial Pons.

Barona Vilar, S. (2019). Inteligencia artificial o la algoritmización de la vida y de la justicia: ¿solución o problema? *Revista Boliviana de Derecho, 28*, 18-49.

Bobadilla, J. (2020). *Machine Learning y deep learning*. *Usando Python, Scikit y Keras* (pp. 14 y ss.). Bogotá: Ediciones de la U. https://books.google.es/books?hl=es&lr=&id=iAAyEAAAQBAJ&oi=fnd&pg=PA11&dq=concepto+de+machine+learning&ots=Qiw2w2pH2t&sig=Rh-J17iFOO_p0bvttpUELVMLwRE#v=onepage&q=concepto%20de%20machine%20learning&f=false

Borges Blázquez, R. (2021). *Inteligencia artificial y proceso penal*. Cizur Menor (Navarra): Ed. Thomson Reuters, Aranzadi.

Boulgouris, N. V. *et al*. (2010). *Biometrics, Theory, Methods, and Applications*. IEEE and Wiley.

Consultora IDC (2018). *The Digitization of the World. From Edge to Core*. https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf (Consultado 15/09/2024).

De Hoyos Sancho, M. (2020). 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*, *76*, octubre-diciembre.

Fariña, R. A y Novo, M. (2002). Heurístico de anclaje en las decisiones judiciales. *Psicothema*, *14*(1).

Kalinowsky, G. (1973). *Introducción a la lógica jurídica*. Editorial Universitaria de Buenos Aires.

Kurzweil, R. (1990). *The Age of Intelligent Machines*. Massachusetts: Ed. MIT Press. 1.ª edición.

Llorente Sánchez-Arjona, M. (2022). Inteligencia artificial, valoración del riesgo y derecho al debido proceso. En S. Calaza López y M. Llorente Sánchez-Arjona (dirs.), *Inteligencia artificial legal y administración de justicia*. Ed. Aranzadi.

Luger, G. F. (2008). *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*. London: Ed. Pearson.

Montesinos García, A. (2022). Justicia penal predictiva. En S Barona Vilar (dir.), *Justicia poliédrica en tiempos de mudanza*. Tirant lo Blanch.

Muñoz Aranguren, A. (2012). Los sesgos cognitivos y el Derecho: el influjo de lo irracional. *El Notario del Siglo XXI*, *42*, marzo-abril. Recuperado el 17/9/2024 de: https://www.elnotario.es/index.php/hemeroteca/revista-42/487-los-sesgos-cognitivos-y-el-derecho-el-influjo-de-lo-irracional-0-53842293707507

Negnevitsky, M. (2011). *Artificial Intelligence: A Guide to Intelligent Systems*. Massachusetts: Addison Wesley. 1.ª edición.

OSCE sobre actividad policial basada en la inteligencia, 2017, p. 6. Recuperado el 9/9/2024 de : https://www.osce.org/files/f/documents/6/4/455536.pd

Perry, W. L., Mcinnis, B., Price, C. C., Smith, S. C. y Hollywood, J. S. (2013). *Predictive Policing. The role of crime forecasting in Law Enforcement operations RAND Corporation*, pp. 33-41. Santa Mónica.

Poole, D. L. y Mackworth, A. K. (2017). *Artificial Intelligence: Foundations of Computational Agents*. Cambridge: Cambridge University Press.

Russell, S. y Norvig, P. (2021). *Artificial Intelligence: A Modern Approach*. London: Ed. Pearson. 4.ª edición.

Schild, U. J. (1998). Criminal Sentencing and Intelligent Decision Support. *Artificial Intelligence and Law*, *6*, 151-202.

Simón, E. y Gaes, G. ASSYST – computer support for guideline sentencing. En *Second International Conference on Artificial Intelligence and Law* (ICAIL-89). Vancouver: ACM Press.

Simón Castellano, P. (2021). Justicia cautelar e inteligencia artificial. *La alternativa a los atávicos heurísticos judiciales*. Barcelona: J. M. Bosch Editor.

Tversky, A. y Kahnemann, D. (sep. 27, 1974). Judgement under uncertainty: Heuristics and Biases. *Science, New Series*, *185*(4157), 1124-1131. Recuperado el 17/03/2024 de: https://www2.psych.ubc.ca/~schaller/Psyc590Readings/TverskyKahneman1974.pdf

Webgrafía

https://www.bbc.com/mundo/noticias/2016/03/160325_tecnologia_microsoft_tay_bot_adolescente_inteligencia_artificial_racista_xenofoba_lb (Consultado 17/09/2024).

https://elpais.com/tecnologia/2018/01/14/actualidad/1515955554_803955.html (Consultado 17/09/2024).

https://www.elmundo.es/motor/2022/06/19/62aeba79fdddff4c408b4588.html (Consultado 17/09/2024).

https://www.psicothema.com/pdf/684.pdf (Consultado: 17/09/2024).

https://www.elnotario.es/index.php/hemeroteca/revista-42/487-los-sesgos-cognitivos-y-el-derecho-el-influjo-de-lo-irracional-o-53842293707507 (Consultado: 17/09/2024).

https://elpais.com/tecnologia/2024-05-28/la-policia-espanola-ya-usa-en-sus-investigaciones-un-sistema-automatico-de-reconocimiento-facial.html (Consultado 29/10/2024).

https://www.clearview.ai/ (Consultado 09/10/2024).

https://www.durham.police.uk/Information-and-advice/Pages/Checlpoint.aspx (Consultado 09/10/2024).

https://www.sdpnoticias.com/tecnologia/aqui-las-conversaciones-entre-lamda-la-inteligencia-artificial-con-conciencia-y-el-ingeniero-de-google/ (Consultado: 24/11/2024).

https://www.abc.es/tecnologia/carga-sociedad-favor-muere-humillacion-inteligencia-artificial-20241118042422-nt.html (Consultado: 18/11/2024).