nº 9 (4º Trimestre 2023)
ISSN: 2792-7962 – ISSNe: 2792-7970

Los sistemas predictivos en la instrucción penal. Una iniciativa docente analítica de las carencias en el camino de un juez autómata

Predictive systems in criminal investigation. A teaching initiative analysing the shortcomings in the path of an automaton judge

Jaime Criado Enguix *Contratado Predoctoral FPU de Derecho Procesal. Universidad de Granada*
https://orcid.org/0000-0003-3894-2610

Cita Sugerida: CRIADO ENGUIX, J. «Los sistemas predictivos en la instrucción penal. Una iniciativa docente analítica de las carencias en el camino de un juez autómata». *Revista Crítica de Relaciones de Trabajo, Laborum*. nº 9 (2023): 77-92.

Resumen

La Inteligencia Artificial (IA) pretende, mediante sistemas programados, emular el pensamiento o la inteligencia humana, y ejercitar acciones como recopilar información, procesarla y validarla, siendo cada vez de mayor aplicación a múltiples áreas. La enseñanza del derecho procesal no ha sido ajena a su influencia. Sería interesante, como propuesta o experiencia piloto, trabajar en una metodología docente en la que, a modo de práctica, el profesor somete el funcionamiento de un determinado algoritmo para la resolución de conflictos, al análisis de sus efectos prácticos por los alumnos. Una vez se conocen los resultados que proporciona la máquina, se invitaría a los alumnos a hacer una disección de lo decidido por la IA, para valorar y evaluar, los sesgos, obstáculos, y problemas que se puedan detectar planteando una comparativa con una posible decisión humana. Esta tarea, a buen seguro, ayudará al alumnado a desarrollar la habilidad de escoger selectivamente la información y el material jurídico del que se dispone, y a razonar, con sólidos argumentos jurídicos, una adecuada solución a los problemas y conflictos jurídicos que se planteen. Y todo ello desde la óptica de la humanización de la tecnología y no de la tecnificación de la decisión.

Palabras clave

Inteligencia artificial; algoritmo predictivo; metodología docente; derecho procesal

Abstract

Artificial Intelligence (AI) aims, through programmed systems, to emulate human thought or intelligence, and to exercise actions such as gathering information, processing it and validating it, and is increasingly being applied to multiple areas. The teaching of procedural law has not been unaffected by its influence. It would be interesting, as a proposal or pilot experience, to work on a teaching methodology in which, by way of practice, the teacher submits the operation of a certain algorithm for conflict resolution to the analysis of its practical effects by the students. Once the results provided by the machine are known, the students would be invited to dissect the decision made by the AI in order to assess and evaluate the biases, obstacles and problems that can be detected by comparing it with a possible human decision. This task will surely help students to develop the ability to selectively choose the information and legal material available, and to reason, with solid legal arguments, an adequate solution to the legal problems and conflicts that arise. And all this from the point of view of the humanisation of technology and not the technification of the decision.

Keywords

Artificial intelligence; predictive algorithm; teaching methodology; procedural law

1. ESTADO DE LA CUESTIÓN. UN ACERCAMIENTO CONCEPTUAL AL SISTEMA DE IA

Asistimos, en los últimos tiempos, a una especie de revolución digital, caracterizada por la sucesión de tecnologías cada vez más sofisticadas e inteligentes, programadas para alcanzar cotas

Fecha Recepción: 28/11/2023 - Fecha Revisión: 07/12/2023- Fecha Aceptación: 14/12/2023
Pags. 77-92

inimaginables, como imitar la inteligencia humana, limitándose, por tanto, no a una mera labor asistencial, sino también, en algunos campos, a una labor decisoria, o sustitutiva de la labor humana.

El ciudadano de a pie, en general, y el jurista, en particular, se ve inmerso en una nebulosa de conceptos normalmente de origen anglosajón, cuyo significado, en muchas ocasiones, desconoce, a saber: *big data*, *cryptocurrencies*, *machine learning*, *blockchain*, *deep learning*, *smart contract*, *data mining*, entre muchos otros.

Todo apunta a una nueva era en la Justicia[1], que no puede coexistir con una Administración de Justicia que aún continúa sobre el papel, y que se sustancia tan sólo de manera presencial, sino que, en los límites que marca la ética y la moral[2], ha de mutar en una justicia moderna, que implemente medios de investigación más eficaces en la prevención del delito, y aumente la eficacia y globalización del proceso.

Se aspira a un cambio en la *praxis* judicial, que convierta el procedimiento en uno ágil, *online* e interconectado, y venga a sustituir el actual modelo ralentizado, escrito y presencial.

No obstante, primero conviene, a los efectos de esta intervención, aclarar, al menos, qué se entiende por Inteligencia Artificial, y, en concreto, entrar a conocer, siquiera sucintamente, su falta de transparencia o explicabilidad, como desafío y aliciente para el alumnado para tratar de comprender y descifrar sus decisiones, y construir las suyas propias.

Esta labor de concienciación y formación del alumnado resulta relevante, por cuanto afirma el Parlamento Europeo[3] que la alfabetización digital sirve para aprovechar plenamente las ventajas de esta tecnología, evitando, con ello, que la falta de conocimientos derive en una brecha digital.

Por consiguiente, no podemos comenzar este análisis sin arrojar, de entrada, el concepto de IA que proporciona el Parlamento Europeo y el Consejo, en su Propuesta de Reglamento por el que se establecen normas armonizadas en materia de Inteligencia Artificial[4]; define, en su art. 3.1, IA como: "*el software que se desarrolla empleando una o varias de las técnicas y estrategias que figuran en el anexo I (a saber, Estrategias de aprendizaje automático, incluidos el aprendizaje supervisado, el no supervisado y el realizado por refuerzo, (...)). Y que puede, para un conjunto determinado de objetivos definidos por seres humanos, generar información de salida como contenidos, predicciones, recomendaciones o decisiones que influyan en los entornos con los que interactúa*".

[1] BARONA VILAR habla de una cuarta revolución industrial (4.0), en que la incorporación de sistemas inteligentes, algoritmos y modelos computacionales en el asesoramiento, predicción y toma de decisiones judiciales es ya una realidad. La autora adopta un interesante enfoque, en que analiza las dudas que plantea la robotización judicial desde el punto de vista de la individualización de la pena, la responsabilidad de la máquina y la afectación de derechos de las personas. *Vid*., Barona Vilar, S.: "Cuarta revolución industrial (4.0) o ciberindustria en el proceso penal: revolución digital, inteligencia artificial y el camino hacia la robotización de la justicia", *Revista Jurídica Digital UANDES*, 2019, p. 5.

[2] Para GÓMEZ COLOMER "*todo tiene sus riesgos. Sostiene que el uso de la IA puede ser muy satisfactorio, muy bueno, muy conveniente, pero también todo lo contrario, de ahí que desde las iniciativas privadas y públicas (...) se incida en definir claramente los límites que la IA no puede traspasar por existir un derecho humano o un derecho fundamental que se opone a ello*". *Vid*. Gómez Colomer, J.L.: "Derechos fundamentales, proceso e inteligencia artificial: una reflexión" en VV.AA.: *Inteligencia Artificial Legal y Administración de Justicia*, Calaza López, S., Llorente Sánchez-Arjona, M. (Dirs.), Ed. Aranzadi, 2022, p. 283.

[3] *Vid*., Parlamento Europeo (2022), Informe sobre la inteligencia artificial en la era digital, Considerando 83º.

[4] Comisión Europea (2021), Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión. Bruselas, 21.4.2021 COM (2021) 206 final. 2021/0106(COD).

En una línea bastante similar, el Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial (*High-Level Expert Group on Artificial Intelligence*) de la Comisión Europea, define IA 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 través 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 (...)*"[5].

Si recurrimos al panorama doctrinal, no podemos hacerlo sin antes citar a MCCARTHY, considerado padre de la IA, quien, en 1956[6], por vez primera, acuño este término, y lo definió como: "*la ciencia e ingeniería de hacer máquinas inteligentes*".

Con el paso del tiempo, la definición de MCCARTHY ha ido cayendo en la obsolescencia. Ya lo advirtió en su momento BARONA VILAR, que, a medida que la tecnología ha ido avanzando, y también lo ha hecho nuestra capacidad de comprensión de la misma, el mismo concepto de IA ha ido variando[7].

MUÑOZ VILLAREAL y GALLEGO CORCHERO[8] ayudan en la concreción de este término. Señalan que, cuando hablamos de IA, fácilmente se puede caer en el error de pensar únicamente en robots, no obstante, estos sistemas abarcan mucho más: están presentes cuando la cámara reconoce los rostros, cuando realizamos búsquedas en Internet, o cuando usamos *Google Maps*, entre otros ejemplos cotidianos.

El campo de la IA no abarca un solo concepto o sistema, sino un conjunto de áreas diversas, que pueden complementar nuestra noción actual de inteligencia artificial, a saber: aprendizaje de máquina, aprendizaje profundo (*Deep learning*), procesamiento de lenguaje natural (PLN), Chat GPT y *chatbots*, entre otros tantos[9].

Frente a ello, no han faltado los autores que han tratado de clasificarlas en compartimentos estancos, entre ellos, sin duda, LÓPEZ DE MÁNTARAS[10]. Este autor distingue entre IA débil e IA fuerte: la primera alude a la ciencia o ingeniería que permite programar ordenadores de forma que realicen tareas que requieren inteligencia; y, la segunda, representa aquella ciencia que permitirá emular la inteligencia humana mediante máquinas. Dos conceptos que aluden, básicamente, a las dos corrientes de la IA, el conexionismo (IA débil) y la ingeniería del conocimiento (IA fuerte). Sin embargo, otros muchos autores, entre ellos, el recién citado, subdividen, a su vez, estas dos categorías

[5] European Commission, Independent High-Level Expert Group on Artificial Intelligence: A definition of AI: main capabilities and disciplines, 8 April 2019, p. 6.

[6] En julio de 1956, se celebró la Conferencia de Verano sobre Inteligencia Artificial en el Dartmouth Collage, organizada por John MaCarthy a la cual asistieron otros nueve noveles científicos, como Marvin Minsky, Claude Shannon, entre otros, para presentar sus respectivos trabajos y poner en común una serie de líneas de investigación para futuras investigaciones. Esta Conferencia constituyó el punto de inflexión a partir del cual arrancaría la IA como campo de investigación.

[7] Barona Vilar, S.: *Algoritmización del derecho y de la justicia. De la Inteligencia Artificial a la Smart Justice*, Ed. Tirant lo Blanch, 2021, p. 94.

[8] Muñoz Villareal, A., Gallego Corchero, V.: "Inteligencia artificial e irrupción de una nueva personalidad en nuestro ordenamiento jurídico ante la imputación de responsabilidad a los robots", *Inteligencia artificial y riesgos cibernéticos. Responsabilidades y aseguramiento*, Ed. Tirant lo Blanch, Valencia, 2019, p. 69.

[9] Al respecto, *vid*., Arenas, M., Arriagada, G., Mendoza, M. y Prieto, C.: *Una breve mirada al estado actual de la Inteligencia Artificial, Pontificia Universidad Católica de Chile*, 2020, dado que despliegan un análisis de rigor en torno a las principales ramas de la IA y sus usos más frecuentes.

[10] López de Mántaras, R.: *Inteligencia Artificial*. Ed. Catarata, 2017, p. 18.

en cuatro modelos. El primero de ellos es el modelo simbólico, luego el conexionista, el evolutivo y el corpóreo.

En dicho concepto, el modelo simbólico se corresponde con la IA que se basa en el razonamiento lógico y en la búsqueda heurística para comprender los problemas, sin la necesidad de integración del sistema a un cuerpo o la necesidad de que esté situado en un entorno real, operando con representaciones abstractas del mundo real, que se basan en lenguaje de representación matemática. Dicho modelo de representación lógica ha avanzado de la simple programación de un sistema hacia nuevas aplicaciones que involucran la toma de decisiones y el aprendizaje.

El aprendizaje permite que la inteligencia artificial pueda realizar procesos de decisión basados en los criterios aprendidos, los cuales pueden ser revisados, o no, por un experto humano. Dicho proceso se programa por medio de algoritmos. El algoritmo, entiende MERCADER UGUINA[11], es "*(…), un conjunto de instrucciones matemáticas, una secuencia de tareas destinada a conseguir un cálculo o un resultado (…), una descripción precisa de los pasos que nos llevan a la solución de un problema planteado*". Por su lado, PERIS MANGUILLOT[12] define algoritmo como "*un conjunto de instrucciones o reglas definidas y no ambiguas, ordenadas y finitas que permite, típicamente, solucionar un problema, realizar un cómputo, procesar datos y llevar a cabo otras tareas o actividades*".

Básicamente, de las definiciones transcritas, se puede convenir que el algoritmo se materializa en un conjunto de instrucciones, comandos o parámetros aplicables a una determinada situación que es el objeto de tratamiento por el algoritmo. Dicho conjunto de soluciones o comandos no es aleatorio, sino que proviene de un número de casos idénticos o similares que permiten al algoritmo basar sus decisiones, en orden a obtener un resultado válido.

Bajo un prisma más científico-técnico, se podría decir, en términos del ingeniero ALBORNOZ CABELLO[13], que el aprendizaje automático (*machine learning*) constituye una de las ramas de la IA que permite incorporar técnicas para que las máquinas aprendan, lo cual se logra "*a través de la búsqueda de algoritmos y heurísticas que trasladan muestras de datos en programas de computadora, sin necesidad de que estos sean totalmente escritos*".

Para interpretar y comprender estos datos, el ser humano no es capaz de programar un conjunto concreto de especificaciones para las tareas que desea emplear la máquina, lo que determina la necesidad de dotar a las computadoras de la capacidad de aprender de la experiencia (pasada) y, por medio de lenguaje matemático e indicadores, llevar a cabo predicciones, correlaciones entre patrones (no guiados por una relación de causalidad[14]), adaptándose con ello a las nuevas situaciones que se les impone, al igual que haría un ser humano, que es donde se aplicaría el aprendizaje automático.

[11] Mercader Uguina, J. R.: *Algoritmos e inteligencia artificial en el derecho digital del trabajo*, Ed. Tirant lo Blanch, 2022, p. 18.

[12] Peris Manguillot, A.: "Algoritmos ¿podemos hacerlos transparentes y trazables en su proceso? en VV.AA.: *Transparencia y explicabilidad de la inteligencia artificial*, Cotino Hueso, L., Castellanos Claramunt, J. (Eds.), Ed. Tirant lo Blanch, Valencia, 2022, p. 71.

[13] Albornoz Cabello, G.A.: *Aplicación del aprendizaje automático supervisado en el mantenimiento predictivo de los motores eléctricos de inducción en las empresas mineras del Perú*, Tesis, Facultad de Ingeniería Eléctrica y Electrónica, Universidad Nacional del Centro del Perú, 2021, pp. 24-25.

[14] Al respecto, COTINO HUESO explica magistralmente que "frente a la contrastación de una hipótesis a partir de los datos, se descubren correlaciones sin conocer previamente la causa. Así sucede al probar casi aleatoriamente la posible correlación entre datos en principio totalmente distantes". *Vid*. Cotino Hueso, L.: "Big data e inteligencia artificial. Una aproximación a su tratamiento jurídico desde los derechos fundamentales", *Dilemata*, núm. 24, 2017, p. 133.

Así, el empleo de la IA resulta cuando menos novedosa, por cuanto tiene la capacidad de autoaprender de experiencias pasadas, analizar el entorno al que se enfrenta, y arrojar predicciones probabilísticas basadas en correlaciones entre patrones, y no en una relación de causalidad.

Esto, no obstante, como bien explica VALLE ESCOLANO[15], "*los datos que alimentan a dichos algoritmos pueden contener sesgos que inciden en la decisión tomada, lo que finalmente tiene impacto en las personas y en sus derechos*".

Un enfoque claro, e interesante para paliar los eventuales efectos adversos de una IA contaminada, sería, sin duda, apostar por su transparencia, en fase de programación (o introducción de los *inputs*), así como en fase de resolución del algoritmo, ya que, lo más complejo es descifrar el íter procesal, o razonamiento argumental, seguido por la máquina para emitir un pronunciamiento/ *output* concreto.

Se sacrifica, ciertamente, el deber de emitir una sentencia fundada en derecho (art. 24.1 CE), en *pro* de un sistema eficaz, y de alto rendimiento. Se plantea un claro dilema. Como bien explica MONEREO PÉREZ[16] los estudiosos de la ciencia jurídica nos vemos forzados a elaborar nuevos conceptos y nuevos derechos, llamados a combatir cualquier peligro, de nuevo cuño, "*por más que, como bien es sabido, el Derecho siempre haya de transitar un paso por detrás de la realidad social que trata de normar*"

Esta labor de combate se pretende ejercer, modestamente, en las aulas, a través de una experiencia docente piloto. Concretamente, vengo impartiendo la asignatura Procesal II (Proceso Penal), en la que se explica, entre otras cuestiones, la relevancia de la tutela cautelar para asegurar el adecuado desarrollo de un proceso, proteger determinados bienes, o proteger a la víctima. Esta labor, que desempeña el Juez en fase de instrucción, puede ser asistida por herramientas inteligentes de justicia predictiva, que se emplean para arrojar una estimación en torno a la probabilidad de reincidencia del *reo*, asesorando al juez en cuanto a la medida cautelar a adoptar que resulte más idónea. No obstante, se vuelve aquí a reiterar, la implementación de la IA en la instrucción penal se enfrenta a serios desafíos y controversias, no sólo por la falta de transparencia algorítmica (donde ahondaremos fundamentalmente), sino por razones morales y éticas, pues la cuestión radica en determinar hasta qué punto deberíamos delegar en una máquina la tarea de decidir cuestiones tan trascendentes como la libertad de la persona.

Por ello, se insiste en las aulas, la apuesta por estas herramientas debe hacerse con suma cautela. La cuestión requiere que se indique, con rigor y claridad, no sólo qué puede y qué no puede hacer una máquina, sin cómo lo puede hacer. A esto precisamente está llamada la metodología docente que se propone.

2. SISTEMA INTELIGENTES DE JUSTICIA PREDICTIVA

Las herramientas inteligentes predictivas constituyen aquellos sistemas tecnológicos que anticipan eventos futuros, y, pueden, por ejemplo, estimar, o predecir, la existencia de posibles y potenciales peligros o riesgos que pueden acontecer, como, *v.gr.*, si un presunto autor del delito - ya detenido - tiene riesgo de reiteración delictiva.

[15] Valle Escolano, R.: "Transparencia en la inteligencia artificial y en el uso de algoritmos" en VV.AA.: *Transparencia y explicabilidad de la inteligencia artificial*, Cotino Hueso, L., Castellanos Claramunt, J. (Eds.), Ed. Tirant lo Blanch, Valencia, 2022, p. 102.

[16] Monereo Pérez, J.L., Rodríguez Escanciano, S. y Rodríguez Iniesta, G.: "Algoritmos e inteligencia artificial. Implicaciones jurídicolaborales: Un enfoque desde la perspectiva de los derechos". *Revista Crítica de Relaciones de Trabajo*, Laborum. nº 8, 2023, p. 11-12.

Los algoritmos no son una ciencia perfecta e infalible, empero sí miden el grado de probabilidad de que algo suceda en base a los datos existentes y recopilados; esta información, recabada en la actividad policial (fase previa al proceso), sin duda, sirve de orientación al Juez para determinar las medidas cautelares a adoptar en fase de instrucción judicial. Así lo argumenta la magistrada SÁNCHEZ CORDERO, en sentencia de la AN (Sala de lo Contencioso), de 30 de septiembre de 2020 (núm. rec. 2187/2019), Fundamento de Derecho Tercero, "*la medición policial del riesgo no es decisiva para el juez, pero es información especializada de asesoramiento útil para la valoración judicial de la «situación objetiva de riesgo para la víctima» que exige la ley procesal en la adopción de medidas cautelares de protección, junto con otros instrumentos de valoración*".

Regularmente, los jueces han basado sus decisiones en las máximas de la experiencia, infiriendo conclusiones a partir de las probabilidades deducidas del conglomerado de los datos y de las circunstancias concretas de cada caso e, inevitablemente, de los resultados de casos anteriores prácticamente iguales o muy similares. Si bien estos profesionales han recibido amplia y extensa formación, y disponen de recursos legales y jurisprudenciales para ello, en ocasiones, con casi total probabilidad, albergan o habrán albergado *"a reasonable doubt"*[17], habida cuenta de que los casos concretos siempre presentan matices, y de que la predicción de comportamientos futuros resulta siempre complicada -máxime para aquellos jueces que no tienen apenas experiencia-.

El hecho de que el Juez o Fiscal en ocasiones albergue una duda razonable no equivale a que la IA se torne en imprescindible para asistir o, en un caso extremo, sustituir la labor de estas autoridades en la toma de decisiones basadas en la predicción de comportamientos futuros y en la detección de posibles riesgos, pues, como apuntábamos, la IA no se trata de una ciencia exacta, sino que presenta sus luces y sus sombras[18]. Su legitimidad en la fase cautelar ha abierto interesantes espacios de debate. De un lado, autores como LLORENTE SÁNCHEZ-ARJONA[19], alertan de los peligros que entraña el que una máquina decida automáticamente la libertad o prisión de una persona (dejando al margen aquellos casos en que se emplea en fase de ejecución, donde la posibilidad de acierto es más alta, al estar respaldada por una sentencia firme de condena); y, de otro, una línea doctrinal conformada, entre otros, por DE HOYOS SANCHO[20], que declaran que, si bien es cierto que resulta complicado admitir juicios predictivos de peligrosidad de un sujeto cuando todavía tiene la presunción de inocencia a su favor, no obstante ello, "*puesto que las medidas cautelares son necesarias y es aceptado que uno de sus fundamentos es evitar el riesgo de reiteración delictiva, no tiene mucho sentido limitar los instrumentos cognitivos del juez*".

Al margen de lo anterior, ya ubicados en fase cautelar, es preciso resaltar que también se ha vertido ingente doctrina en torno a si la labor predictiva del algoritmo debe ser asistencial, complementaria o sustitutoria. Más allá de cualquier posicionamiento, lo cierto es que, a día de hoy, existen cuestiones pendientes por resolver, como: ¿quién responde de los perjuicios ocasionados por

[17] Por fortuna, nuestro ordenamiento jurídico prevé un sistema de doble instancia, un sistema de recursos a través del cual las partes pueden procurar la modificación o anulación de una resolución judicial aún no firme que les perjudica.

[18] Además, ésta - deshumanizar la justicia - no ha de ser el fin primordial de la IA, sino más bien, como indica la jueza e investigadora CUATRECASAS MONFORTE - "*dotar a la Administración Pública de mejores y más medios técnicos que sirvan para auxiliar y dar soporte a los ya existentes y a las personas que están al servicio de la investigación de las causas penales*". *Vid*., Cuatrecasas Monforte, C.: *La Inteligencia Artificial en el proceso de instrucción español: posibles beneficios y potenciales riesgos*, Universita Ramon Llull, 2022, p. 316.

[19] Llorente Sánchez-Arjona, M.: "Inteligencia artificial, valoración del riesgo y derecho al debido proceso" en VV.AA.: *Inteligencia artificial legal y administración de justicia*, Calaza López, S., Llorente Sánchez-Arjona, M. (Dirs.), Ed. Aranzadi, 2022, p. 383.

[20] 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*, núm. 76, octubre-diciembre 2020, p. 32.

un algoritmo predictivo que arroja una estimación errónea? ¿cómo rebatir un resultado algorítmico, cuyo razonamiento decisorio es opaco, o inaccesible al justiciable?

Personalmente, hasta tanto se diluciden estas cuestiones (que escapan del objeto de esta investigación), lo prudente es, como bien establece MARTÍN DIZ[21], "*posicionarse hacia la admisibilidad de sus funciones asistenciales de apoyo a decisiones jurisdiccionales (o arbitrales), dando soporte a las mismas desde las recomendaciones u orientaciones que puedan elaborar, pero sin sustituir ni reemplazar la actividad jurisdiccional o arbitral humana*".

En el marco del proceso penal, son varios los algoritmos predictivos de riesgo que se utilizan previamente por la policía judicial, auxiliando en base a la estimación arrojada por el algoritmo al Juez en la toma de decisiones concretas, que afectan a la libertad del presunto delincuente, como la adopción de medidas cautelares, o medidas con posteridad al juicio, como las relativas a la libertad condicional.

De entre las distintas herramientas que integran la "justicia predictiva", cabe comenzar por *Correctional Offender Management Profiling for Alternative Sanctions* (en adelante, COMPAS) técnica de evaluación de riesgos, por excelencia, empleada en el ámbito judicial penal estadounidense, desarrollada en 1998 por una empresa privada (*Northpointe*). Contiene un algoritmo que, por un lado, calcula la probabilidad de que un acusado reincida y, por otro, sugiere qué tipo de supervisión/tratamiento debería este recibir en prisión, y todo ello a partir de la información recopilada sobre su conducta pasada y la realización de un cuestionario. No obstante, el origen de los datos, la totalidad de los factores que se tienen en consideración y el peso que se otorga a cada uno de ellos, permanecen ocultos, por lo que resulta imposible entender qué hay detrás de los resultados que el sistema arroja.

Además de COMPAS, existen otras herramientas inteligentes de evaluación de riesgos de reincidencia, entre los que cabe destacar:

— El algoritmo LSI-R (*Level of Service Inventory-Revised*), se usa en los permisos de salida y la libertad condicional de los procesados; aquí, la ponderación utiliza criterios como los antecedentes penales, lugar de residencia, educación, empleo, ocio, familia, problemas de alcohol o drogas, actitudes emocionales y personales.

— Por otro lado, se encuentra el algoritmo HART (*Harm Assessment Risk Tool*), diseñado por la policía de Durham (Inglaterra), cuya función es predecir el nivel de riesgo de los individuos de cometer delitos en un lapso de dos años, en base a criterios como la edad, el sexo o el historial de delitos; este algoritmo ayuda a determinar si el individuo que tiene un riesgo bajo puede ser enviado por los agentes de policía a un programa de rehabilitación denominado *Checkpoint*.

— En España, contamos, entre otros, con experiencias constatables mediante el sistema *VioGen*. Esta es una técnica informática de ámbito nacional, que emana de la Secretaría de estado de seguridad del Ministerio del Interior, en funcionamiento desde 2007, aunque ha sido objetiva de numerosas actualizaciones. Entre sus cometidos destaca la realización de valoraciones policiales del riesgo de las víctimas denunciantes de sufrir una nueva agresión o acto de violencia de género, y, en función del resultado, protegerlas.

[21] *Vid*., Martín Diz, F.: "Modelos de aplicación de inteligencia artificial en justicia: asistencial o predictiva versus decisoria", *Justicia algorítmica y neuroderecho: Una mirada multidisciplinar*, Ed. Tirant lo Blanch, Valencia, 2021, p. 83.

En síntesis, los diversos tipos de algoritmos de riesgo analizados ponderan diferentes variables y son utilizados primordialmente con el objetivo de anticipar la decisión criminal o la reincidencia, e inclusive apoyan la toma de decisiones relacionadas con la libertad, rehabilitación o servicios sociales.

Tales técnicas plantean claros beneficios y utilidades tanto en el ámbito policial, como judicial y, por supuesto, en el ámbito académico, donde nos detendremos:

— En el ámbito jurisdiccional propiamente dicho, puede no sólo mejorar enormemente la eficiencia de los trámites judiciales, y mejorar la calidad de las decisiones emitidas por jueces, sino que les permite observar de manera estadística los precedentes más importantes relacionados a un caso, y muchas de sus variables, por lo que están llamados a convertirse en un asistente clave de los jueces.

— En la actividad policial, la capacidad de anticipar un futuro delito refuerza la labor policial predictiva de la policía, y le permite una asignación de recursos más eficiente.

— En el plano académico también está llamado a reportar utilidades, principalmente por el desafío que supondría plantear al alumnado la tarea de descifrar, con respuestas jurídicas, el razonamiento o iter argumental seguido por el algoritmo.

Como bien enfatiza ESPARZA LEIBAR[22], "*los riesgos que deben ser conjurados para generar confianza se centran en el respeto a los derechos fundamentales, para lo que cuestiones como la transparencia, la responsabilidad, la privacidad o los sesgos en la adopción de decisiones autónomas, deben ser correctamente detectadas y resueltas*".

En nuestra propuesta piloto, se trabajará sobre el valor de la transparencia, como alternativa clara para paliar la opacidad del algoritmo, y detectar los sesgos concurrentes.

Los algoritmos no pueden argumentar el porqué de sus decisiones. Ello resulta cuando menos objetable, por cuanto desprovee al justiciable de una base sobre la que recurrir, vulnerando su derecho a un recurso efectivo, que forma parte del contenido del derecho a la tutela judicial efectiva (art. 24 CE). Por ello, señala BARONA VILAR[23], resulta preciso "*construir una IA explicada*". Sería interesante iniciar esta labor desde las instancias académicas, en las aulas.

3. LA INEXPLICABILIDAD DEL ALGORITMO

Un foco de riesgo, así como de suspicacia respecto del algoritmo, reside en la falta de transparencia, la difícil explicabilidad, y la ausencia de conocimiento de la regla algorítmica concreta que se ha empleado para basar una decisión judicial.

Es común que los algoritmos funcionen siguiendo un modelo de caja negra, es decir, el usuario tras introducir los datos en el modelo obtiene unos resultados, pero se desconoce el camino o íter recorrido para llegar a los mismos.

En términos de BATHAEE, el principal problema de la caja negra o black box puede ser definido como la "*incapacidad de comprender plenamente el proceso de toma de decisiones de un*

[22] Esparza Leibar, I.: "La inteligencia artificial y el derecho fundamental a la protección de datos de carácter personal" en VV.AA.: *Justicia algorítmica y neuroderecho…op.cit.,* p. 276.

[23] Barona Vilar, S.: "Una justicia "digital" y "algorítmica" para una sociedad en estado de mudanza", *Justicia algorítmica y neuroderecho…op.cit.*, p. 54.

sistema de IA y la incapacidad de predecir las decisiones o los resultados de la mismos"[24]. Explica CATALINA BENAVENTE[25] que el peligro en torno a la utilización de Inteligencia Artificial aumenta cuando se utilizan algoritmos de "caja negra", que impiden en base a criterios la máquina ha decidido o seleccionado.

Las resoluciones basadas en algoritmos adolecen de una clara falta de transparencia que obedecen a diversos factores, en este punto, concretamente, nos referiremos a dos: de un lado, a la complejidad de la estructura de un algoritmo, como en el caso de una red neuronal profunda, en que existen miles de neuronas artificiales que trabajan juntas de modo difuso para resolver un problema; y, de otro, a la difícil explicabilidad del sistema predictivo, esto es, a la dificultad para traducir en un lenguaje fácil y comprensible el íter seguido por la máquina predictiva para adoptar una decisión.

Tal opacidad o difícil explicabilidad del sistema de IA puede lesionar el derecho de defensa[26], pues si se priva a los litigantes de información acerca de la regla algorítmica empleada en la toma de decisiones, el justiciable desconoce la motivación o fundamentación de la misma, lo que le desprovee de una base concreta sobre la que recurrir o impugnar, pilar básico del derecho a la tutela judicial efectiva en el proceso penal (art. 24 CE).

Una vez se accede al código fuente, es posible detectar sesgos, o tendencias que favorecen a un colectivo frente a otro. Y es que, la IA, en relación al proceso penal, suscita serias dudas no sólo con relación al cumplimiento del principio de transparencia, sino también al de no discriminación. Y esto último trae causa de los sesgos discriminatorios. Riesgos estos que, a juicio de la Comisión Europea, tienen la consideración de alto riesgo[27], al tener un impacto negativo en los derechos fundamentales de las personas. Nos detendremos en los sesgos que arroja el algoritmo.

4. TENDENCIA AL SESGO ALGORÍTMICO

Un riesgo crítico y propio de operar con algoritmos en sede de justicia es, sin duda, la concurrencia de sesgos. Ya los reconocidos psicólogos israelís, KANHEMAN y TVERSKY, se embarcaron, en su investigación, en el análisis de la teoría relativa a que las decisiones del ser humano, *per se*, no son completamente objetivas, ni informadas. Mantienen que el ser humano dispone de información parcial, a la que hay que sumar la experiencia, las creencias, los valores y los prejuicios anteriores, todo lo cual, inconscientemente, le infunde una determinada visión, prejuiciosa, o sesgada, de la realidad, con determinadas inclinaciones o preferencias. Ello, unido al hecho de que se presentan al ser humano abundantes contextos en los que tiene que adoptar una decisión, le lleva, inexorablemente, a mecanizar sus decisiones, ya sesgadas. Esto, señalan los referidos autores, es lo que se conoce como los heurísticos (o "atajos") de pensamiento[28].

El razonamiento humano, no es tan perfecto, ni meditado, como pudiera pensarse, más bien al contrario, cae, cada vez más, en automatismos, similares a los de una máquina. NIEVA FENOLL[29]

[24] Bathaee, Y.: "The Artificial Intelligence Black Box and the Failure of Intent and Causation". *Harvard Journal of Law & Technology*, vol. 31 (2), 2018, pp. 893 y ss.

[25] Catalina Benavente, Mª. A._ "La identificación de pasajeros "sospechosos" a través del tratamiento masivo de sus datos PNR" en VV.AA.: *Justicia algorítmica y neuroderecho…op.cit.,* p. 352.

[26] En este sentido, NIEVA FENOLL apunta que no es posible elaborar una mínima estrategia defensiva si no se puede conocer cómo decide la máquina. *Vid*. Nieva Fenoll, J.: *Inteligencia artificial y proceso judicial*, Ed. Marcial Pons, Madrid, 2018, pp. 139 y ss.

[27] Comisión Europea: "*Nuevas normas sobre la Inteligencia Artificial: preguntas y respuestas*", Bruselas, 21 de abril de 2021, p. 2.

[28] Kanheman, D., Slovic, P. y Tversky, A.: "Judgment under Uncertainty: Heuristics and Biases", *Cambridge University Press*, 1982, pp. 3 y ss.

[29] Nieva Fenoll, J.: *Inteligencia artificial y proceso judicial,* Ed. Marcial Pons, 2018, p. 45.

profundiza de manera excelente en esta cuestión, y explica que un heurístico "*es una especie de directriz general que podemos seguir los seres humanos para tomar una decisión. A la hora de decidir, cualquier persona acostumbra a recordar una situación análoga en la que obró de una determinada forma, determinando así el patrón de conducta más frecuente, volviendo a actuar en la situación concreta del mismo modo*". Por tanto, las situaciones se generalizan, se desvanecen los matices, y se toman decisiones automáticas en espacios muy breves de tiempo.

Todo lo dicho nos reafirma básicamente como seres humanos, que, fruto del entorno social, cultural, económico e ideológico en el que nos desenvolvemos, y presos de un ritmo de vida cada vez más acelerado, nos volvemos más propensos a esos heurísticos de pensamiento, que no hacen sino automatizar decisiones sesgadas.

Por otro lado, si nos centramos ya en el término sesgo, no podemos comenzar sin antes recurrir a la definición que acuñan los expertos psicólogos americanos DANKS y LONDON[30], que entienden que se trata de: "*a deviation from a standard*", es decir, una desviación de un estándar. En función de la norma que se tome como punto de referencia, hablaremos de un tipo de sesgo, o de otro: verbigracia, estaremos ante un sesgo estadístico cuando una estimación del algoritmo se aleja de una norma, o de un valor real estadístico; o, ante un sesgo de tipo moral, cuando el algoritmo está nutrido o entrenado con ciertas convicciones que se desvían de una norma moral, etc.

Desde Europa, las instituciones entienden este término de un modo similar. El Parlamento Europeo, en Resolución de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas, en el apartado definiciones, art. 4.l), define sesgo como "*toda percepción personal o social prejuiciosa de una persona o de un grupo de personas sobre la base de sus características personales*".

De lo expuesto se puede extraer una primera idea en claro, y es que este pensamiento sesgado, es, en exclusiva, consecuencia de la condición humana. Y es éste el que, mediante ingeniería informática, traslada estos sesgos a las máquinas inteligentes que crea, y entrena, a su libre albedrío. Recuérdese que el algoritmo, *per se*, es neutro, cuestión distinta es el patrón que le imprime su creador, que, en la mayoría de las ocasiones, no es otra cosa que su pensamiento sesgado.

Por tanto, el sesgo algorítmico proviene, básicamente, del reflejo en él de estos heurísticos de pensamiento de su programador - cargado de prejuicios, tendencias e inclinaciones sociales, morales o culturales -. Este, a nuestro juicio, es el orden lógico que siguen los sesgos algorítmicos: primero, el programador crea el algoritmo, al que imprime un patrón discriminatorio inicial; y, el algoritmo, como *software* o aplicación informática inteligente, que presumimos *iuris tantum* funciona correctamente, sin disfunciones, se programa para que automatice, aprenda y perpetúe el patrón en cuestión[31].

[30] Danks, D., London, A.L.: "Algorithmic Bias in Autonomous Systems", *Proceedings of the 26thInternational Joint Conference on Artificial Intelligence,* 2017, p. 4692.

[31] MARTÍN LÓPEZ aporta otro enfoque en relación a esta secuencia, y apoya la tesis de que todo proceso decisional, ya sea computacional o humano, por su propia naturaleza, o reglas de funcionamiento, incorpora sesgos. "*(...) su función intrínseca es, precisamente, la de seleccionar y priorizar unos datos o inputs frente a otros, por lo que los criterios utilizados para ello también mostrarán determinadas inclinaciones. Por tanto, no parece posible una decisión humana o computacional carente de cualquier tipo de sesgo, siempre que aquélla no sea aleatoria*". *Vid*., Martín López, J.: "Inteligencia artificial, sesgos y no discriminación en el ámbito de la inspección tributaria" en *Artículos científicos, crónica tributaria*, núm. 182, 2022, p. 66.

Sin ánimo de ser exhaustivos, cabe señalar, mínimamente, que existen diferentes tipos de sesgos. GUTIÉRREZ DAVID[32], realiza la siguiente clasificación:

i) *preexistentes*, esto es, que existen desde el momento de la génesis del algoritmo, que se ha sido entrenado con datos sesgados;

ii) *técnicos*, es decir, aquellos que surgen por anomalías o disfunciones técnicas del sistema;

iii) o *emergentes*, aquellos que surgen y sólo son detectables después de que los usuarios hayan interactuado con el algoritmo.

En atención al matiz arrojado *ut supra*, cabría descartar por nuestra parte los sesgos técnicos, o dimanantes de una actuación del algoritmo anómala; ello porque entendemos que el proceso computacional es neutral y objetivo hasta que se demuestre lo contrario. Destruida, en su caso, esta presunción, ya sí, cabría hablar y admitir que, a los sesgos humanos iniciales y preexistentes en el algoritmo, se le han adicionado sesgos o desviaciones generadas por un desorden o alteración del funcionamiento interno del algoritmo.

Dicho lo cual, parece que, salvo en el caso de sesgos técnicos - que pueden ser salvables mediante una correcta intervención periódica por parte de un equipo de informáticos especializado - en el resto de supuestos, existe un elemento psicológico, propenso al sesgo cognitivo, que el ser humano, consciente, o no, imprime en el patrón del algoritmo, y que resulta indisociable de su código fuente. Por tanto: ¿cómo evitar el sesgo en el algoritmo? Pues bien, la cuestión no radica tanto en tratar de crear una IA completamente libre de sesgos, pues, como venimos señalando, el algoritmo y los sesgos son, en la mayoría de ocasiones, indisociables, por la condición humana de su creador. La cuestión radica en no confundir el término sesgo, con el término discriminación jurídica. Este último, según el art. 1 del Convenio de la OIT, "*(...) comprende: a) cualquier distinción, exclusión o preferencia basada en motivos de raza, color, sexo, religión, opinión política, ascendencia nacional u origen social, que tenga por efecto anular o alterar la igualdad de oportunidades o de trato en el empleo y la ocupación (...)*". El art. 14 CE establece que "*los españoles son iguales ante la ley, sin que pueda prevalecer discriminación alguna por razón de nacimiento, raza, sexo, religión, opinión o cualquier otra condición o circunstancia personal o social*". Por tanto, entramos en terreno de discriminación jurídica cuando una resolución judicial favorece a un colectivo frente a otro, por una preferencia basada en los motivos o causas sensibles prevista en la ley: religión, raza, color, etc., y ello se detecta.

El sesgo no conlleva necesariamente una discriminación jurídica. Habría que atender cada caso específico, y analizar, cuidadosamente, si el trato diverso proporcionado a un justiciable descansa o no en alguna de las causas protegidas, o más bien responde a un parámetro objetivo.

[32] Gutierrez David, M.E.: "Administraciones inteligentes y acceso al código fuente y los algoritmos públicos. Conjurando riesgos de las cajas negras decisionales", Derecom, 30, 2021, p. 167. Por su lado, el profesor AMUNÁTEGUI ha identificado tres tipos de sesgos que puede adquirir un sistema de IA: i) el sesgo de *interacción*, que se presenta cuando el propio programador, inadvertidamente, introduce un sesgo por el modo en el que interactúa con él; ii) el sesgo *latente*, aparece por un error técnico de la máquina, cuando efectúa correlaciones inapropiadas entre los datos, dando lugar a falsos nexos entre los datos; iii) y sesgo de *selección*, el cual tiene lugar cuando los datos recopilados no es suficientemente representativo de la diversidad del medio social. *Vid*., Amunátegui Perelló, C. y Madrid, R.: "Sesgo e Inferencia en Redes Neuronales ante el Derecho", en VV.AA.: *Inteligencia Artificial en América Latina y el Caribe. Ética, Gobernanza y Políticas*, Aguerre, C. (Ed.), Bs. As., CETyS Universidad de San Andrés, 2020, p. 5

Relacionado con la posible discriminación jurídica de un algoritmo, cabe señalar que los efectos que derivan de ello resultan más nocivos o adversos en el ámbito de la IA que, en el ámbito humano, al manejar estos sistemas de *big data* y *machine learning* un gran volumen de datos, que les permite perpetuar el trato discriminatorio a gran escala, si no se interviene o ejerce sobre ellos el debido control *ex ante*.

En definitiva, no es posible una absoluta erradicación del sesgo; como explica OLIVER CUELLO[33], con cita a GONZÁLEZ DE FRUTOS, incluso en caso de algoritmos completamente libres de sesgos, "*la discriminación se puede producir cuando el sistema aprende de los datos históricos. Es el problema de predecir el futuro con base en hechos pasados: si la discriminación está en las series históricas, el algoritmo va a interiorizarla mediante el aprendizaje automático*". Por tanto, consideramos, los esfuerzos se habrían de invertir en analizar si el sesgo concurrente tiene un impacto discriminatorio negativo, por favorecer a un colectivo frente a otro en situación similar, por cualesquiera de las causas especiales protegidas en la CE, a saber: género, raza, religión, entre otras. Detectado, en su caso, el sesgo discriminatorio, procedería su inmediata corrección para evitar una amplificación de los perjuicios.

Lógicamente, no podemos entrar a analizar la naturaleza discriminatoria de una medida cautelar, si antes no hemos podido acceder al código fuente, o, al menos, hemos comprendido su fundamentación. Es necesaria una revisión *ex ante* del sistema de Inteligencia Artificial, en especial si se trata de algoritmos que afectan a la administración de justicia. Por lo tanto, a la resolución de este segundo problema, le ha de preceder necesariamente la resolución del primer problema analizado, su falta de transparencia.

Frente a los riesgos que plantea la IA, la verdadera cuestión, propone PLANCHADELL-GARGALLO[34], está en plantearnos si tiene más relevancia, o peso, las ganancias económicas de la empresa que genera y gestiona el algoritmo, o los derechos de la persona afectada por el mismo. Opino, en el mismo sentido que la autora, que la respuesta es más que evidente; por ello, a título personal, en las aulas, se insiste en la falta de transparencia del algoritmo, no sólo como primer paso para detectar sesgos, sino también como oportunidad docente para incidir en la necesaria motivación de las resoluciones judiciales, como garantía del derecho de defensa del justiciable.

5. DESCRIPCIÓN DE LA INICIATIVA

En el grado en Derecho, en la Universidad de Granada, imparto, entre otras asignaturas, la asignatura Derecho Procesal Penal, para alumnos de tercero de carrera. Esta asignatura, al menos en mi experiencia como alumno, por su complejidad terminológica y extensión de normas técnicas, puede antojarse poco atractiva de estudiar, a lo que hay que adicionar la breve duración de un cuatrimestre.

La enseñanza de esta materia viene reclamando, sin duda, un nuevo enfoque metodológico. La propuesta piloto que se ofrece se centraría, concretamente, en una fase concreta de este proceso jurisdiccional: la fase de instrucción. Veamos en qué consiste:

En primer lugar, los alumnos recibirían una lección teórica acerca de la fase de instrucción, a saber, entre otros extremos: i) la naturaleza jurídica de la instrucción; ii) órganos competentes; iii)

[33] Oliver Cuello, R.: "Big Data e inteligencia artificial en la Administración Tributaria" en *Revista de Internet, Derecho y Política*, núm. 33, octubre 2021, p. 10, que cita a González de Frutos, U.: "Inteligencia artificial y Administración tributaria" en VV.AA.: *Fiscalidad e inteligencia artificial: Administración tributaria y contribuyentes en la era digital*, Serrano Antón, F. (Dir.), Ed. Aranzadi, Pamplona, 2020, p. 157.

[34] Planchadell-Gargallo, A.: "Inteligencia artificial y medidas cautelares" en VV.AA.: *Justicia algorítmica y neuroderecho…op.cit.*, p. 409.

iniciación del procedimiento; iv) desarrollo del sumario y diligencias de investigación; v) medidas cautelares; vi) garantías procesales del detenido; viii) y fin de la instrucción.

Tras una asimilación teórica de los preceptos, y superados los correspondientes *tests* o pruebas en clase para analizar el grado de comprensión de los conceptos teóricos, pasaríamos a la acción práctica, de indudable relevancia en el ámbito jurídico.

En un aula, por ejemplo, de setenta alumnos, se pediría a estos que se organizasen en grupos de cuatro o cinco. A cada grupo se le asignaría la resolución de varios supuestos prácticos "reales", que cuentan, como es lógico, con la decisión del Juez que en su momento fue competente para conocer del asunto.

Se detallaría a cada grupo de trabajo los datos que rodean el caso: la identidad del presunto infractor, sus antecedentes, características, las circunstancias agravantes o atenuantes concurrentes, así como el resto de indicadores que deberán valorarse: como la naturaleza o gravedad del hecho, la situación familiar y laboral del sujeto, entre otros factores.

En base a estos datos, el alumno deberá decidir y motivar adecuadamente, conforme a las clases teóricas recibidas, si adopta o no una medida cautelar personal. Se les advertirá previamente en la correspondiente lección teórica que tales medidas cautelares son sin duda las más trascendentes, habida cuenta de los bienes jurídicos a los que afectan, ya que suponen una limitación o una prohibición en el ejercicio de libertades individuales, señaladamente de la libertad de movimientos, de la que se puede privar al sujeto a través de la detención o de la prisión provisional, o se puede limitar su ejercicio con la libertad provisional. Medidas cautelares que por representar intromisiones de muy diferente entidad y, en la medida en que afectan a libertades fundamentales, han de acordarse sólo cuando resulte estrictamente necesario para el proceso penal.

Cada grupo de trabajo deberá decidir y razonar, en base al supuesto planteado, por ejemplo, si procede o no a decretar prisión provisional, en cualquiera de sus tres modalidades, comunicada, incomunicada o atenuada, o si opta por la imposición de un programa de rehabilitación, o de una medida de seguridad, u opta por la libertad provisional, o la detención del reo.

Evidentemente, aunque tales medidas persigan un mismo objetivo, el modo en que se produce la privación en cada una de ellas es diferente. Por ello, independientemente de la opción por la que se opte, el grupo deberá, en su auto correspondiente, como mínimo, reproducir los hechos, la solicitud de medida cautelar, y los razonamientos jurídicos que los lleva a acordar o rechazar la petición/solicitud planteada, sin perjuicio de que uno o más miembros del grupo tenga una opinión discrepante con respecto al resto, y pueda hacer valer su punto de vista a la clase. Subráyese que la finalidad del trabajo en grupo no es tanto que todos los miembros del equipo alcancen una solución idéntica, como que entablen un diálogo sano, expongan sus posturas, y, a través de la escucha y el debate, se retroalimenten y puedan sacar sus propias conclusiones, y decidir finalmente si mantienen la misma postura, deciden matizarla, o cambiarla de plano.

El debate previo entre los compañeros, al darse en un contexto más distendido y cerrado, y al contar con varios puntos de vista, a buen seguro, les ayudará a aprender y extraer conclusiones oportunas que ayuden a reforzar su postura o, incluso, a cambiarla. Se valorará especialmente que la decisión que plantee cada grupo se corresponda con una adecuada argumentación en torno a la concurrencia o no de los presupuestos para que se pueda decretar, por ejemplo, la prisión provisional.

De forma simultánea, el docente someterá el funcionamiento de un determinado algoritmo para estimación del riesgo de reincidencia, al análisis de sus efectos prácticos por los alumnos. Consistiría en utilizar - previo acuerdo con la empresa o entidad que ostenta la patente - por ejemplo, el algoritmo PSA (*Public Safety Assessment*) - el cual se viene usando para determinar si el procesado puede permanecer en libertad hasta la etapa del juicio, teniendo en cuenta el nivel de riesgo de reincidencia o de no comparecencia a la audiencia de juicio - al objeto de someter a su resolución y criterio los casos planteados en clase. Una vez se conocen los resultados (*output*) que proporciona la máquina, y la decisión algorítmica por la que opta, se invitaría a cada grupo de cuatro a hacer una exhaustiva disección de lo decidido por la IA, para valorar y evaluar los sesgos, obstáculos, y problemas que se puedan detectar planteando una comparativa con su decisión, y también confrontándola con la que puede plantear el docente al respecto.

En este punto, se les pediría a los alumnos que argumentasen, por escrito, en qué medida la decisión o tutela cautelar adoptada por la máquina se ajusta a derecho, resulta o no respetuosa con los derechos fundamentales (como el derecho a la presunción de inocencia) y, sobre todo, se les pedirá que justifiquen si a su criterio, la máquina valora o no adecuadamente la concurrencia de los presupuestos de procedibilidad de una medida cautelar: *periculum in mora*, *fumus boni iuris*, proporcionalidad, fin legítimo, etc.

Esta fase argumentativa última de la actividad práctica es, a mi juicio, la más relevante, pues incita a los alumnos no sólo a analizar, estudiar y repasar los requisitos de adopción de una medida cautelar, y a manejar los preceptos de la LECrim vinculados a esta materia (Libro II, Título VI LECrim), sino a llevar a cabo un notorio y excepcional esfuerzo argumentativo para, de un lado, motivar su decisión "judicial" y, de otro, contradecir o corroborar la decisión cautelar adoptada por el algoritmo (y lo que es más importante en este segundo aspecto, descifrar no tanto la decisión que adopta la máquina, si no el por qué la adopta).

Esto plantea al alumnado un "reto" estimulante, en el que se le incita a confrontar su decisión con la del robot, y a reunir los fundamentos jurídicos adecuados para apoyar su postura, y contrarrestar, o corroborar otra. Esta tarea, de descifrar y justificar la solución algorítmica, y de defender la suya propia, a buen seguro, ayudará al alumnado a desarrollar esa IA *explicada* que acuña BARONA VILAR, y les ayudará a desarrollar, desde su fase formativa, la habilidad de escoger selectivamente la información y el material jurídico del que se dispone, y a razonar, con sólidos argumentos jurídicos, una adecuada solución a los problemas y conflictos jurídicos que se planteen. Y todo ello desde la óptica de la humanización de la tecnología y no de la tecnificación de la decisión.

6. RESULTADOS PREVISTOS Y VALORACIÓN FINAL

Ciertamente, las aulas, por la extensión de alumnos - que dificulta una atención personalizada -, la complejidad de términos, y la concentración de sesiones diariamente, suscitan un bajo ánimo en el alumnado, que les impide afrontar con solvencia y claridad la concreta asignatura que imparto. Hablamos, recuérdese, de una disciplina de suma relevancia para la formación de un jurista, por cuanto le permite conocer y dominar las herramientas para representar y hacer valer los derechos subjetivos del justiciable, en los cauces jurídicos previstos en la ley. Como posibles ejercientes, es de obligado dominar la legislación procesal, en cualesquiera de sus vertientes, para comprender el funcionamiento de los tribunales, y los mecanismos de defensa de que dispone todo ciudadano. Este aprendizaje ha de comenzar en las aulas.

Con la metodología que se propone, se espera una participación más activa por parte del alumnado, no sólo por implicar el uso de novedosos sistemas inteligentes, algo con lo que parece estar

muy familiarizados las nuevas generaciones, sino también por el reto que se plantea, que es tratar de escudriñar, y descifrar, el funcionamiento del algoritmo.

Se reputará que esta metodología ha tenido éxito si se aprecia no sólo más intervenciones en clase, sino también un avance cualitativo en el dominio de aspectos básicos del proceso judicial penal.

A modo de conclusión, consideramos que la enseñanza del Derecho Procesal debe beneficiarse de los nuevos avances tecnológicos, no sólo como alternativa para amenizar las clases - sin perder nunca de vista el rigor científico-jurídico - sino también como preparación para los nuevos desafíos del mañana, en que el alumnado, ya en calidad de abogado, o de representado, a buen seguro se verá en la situación de recurrir una decisión cautelar asistida por un algoritmo predictivo, cuyo *output* no comparta, y le cause un gravamen específico. Es preciso preparar a las nuevas generaciones con los recursos propios de su tiempo.

7. BIBLIOGRAFÍA

Albornoz Cabello, G.A.: *Aplicación del aprendizaje automático supervisado en el mantenimiento predictivo de los motores eléctricos de inducción en las empresas mineras del Perú*, Tesis, Facultad de Ingeniería Eléctrica y Electrónica, Universidad Nacional del Centro del Perú, 2021.

Amunátegui Perelló, C. y Madrid, R.: "Sesgo e Inferencia en Redes Neuronales ante el Derecho", en VV.AA.: *Inteligencia Artificial en América Latina y el Caribe. Ética, Gobernanza y Políticas*, Aguerre, C. (Ed.), Bs. As., CETyS Universidad de San Andrés, 2020.

Arenas, M., Arriagada, G., Mendoza, M. y Prieto, C.: *Una breve mirada al estado actual de la Inteligencia Artificial, Pontificia Universidad Católica de Chile*, 2020.

Barona Vilar, S.: "Cuarta revolución industrial (4.0) o ciberindustria en el proceso penal: revolución digital, inteligencia artificial y el camino hacia la robotización de la justicia", *Revista Jurídica Digital UANDES*, 2019.

Barona Vilar, S.: "Una justicia "digital" y "algorítmica" para una sociedad en estado de mudanza", en VV.AA.: *Justicia algorítmica y neuroderecho. Una mirada multidisciplinar,* Barona Vilar, S. (Ed.), Ed. Tirant lo Blanch, 2022.

Barona Vilar, S.: *Algoritmización del derecho y de la justicia. De la Inteligencia Artificial a la Smart Justice*, Ed. Tirant lo Blanch, 2021.

Bathaee, Y.: "The Artificial Intelligence Black Box and the Failure of Intent and Causation", *Harvard Journal of Law & Technology*, vol. 31, núm. 2, 2018.

Catalina Benavente, Mª. A.: "La identificación de pasajeros "sospechosos" a través del tratamiento masivo de sus datos PNR" en VV.AA.: *Justicia algorítmica y neuroderecho. Una mirada multidisciplinar,* Barona Vilar, S. (Ed.), Ed. Tirant lo Blanch, 2022.

Cotino Hueso, L.: "Big data e inteligencia artificial. Una aproximación a su tratamiento jurídico desde los derechos fundamentales", *Dilemata*, núm. 24, 2017.

Cuatrecasas Monforte, C.: *La Inteligencia Artificial en el proceso de instrucción español: posibles beneficios y potenciales riesgos*, Universita Ramon Llull, 2022.

Danks, D., London, A. L.: "Algorithmic Bias in Autonomous Systems", *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, 2017.

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*, núm. 76, octubre-diciembre 2020.

Esparza Leibar, I.: "La inteligencia artificial y el derecho fundamental a la protección de datos de carácter personal" en VV.AA.: *Justicia algorítmica y neuroderecho. Una mirada multidisciplinar*, .: BARONA VILAR, S. (Ed.), Ed. Tirant lo Blanch, Valencia, 2021.

Gómez Colomer, J.L.: "Derechos fundamentales, proceso e inteligencia artificial: una reflexión" en VV.AA.: *Inteligencia Artificial Legal y Administración de Justicia*, Calaza López, S., Llorente Sánchez-Arjona, M. (Dirs.), Ed. Aranzadi, 2022.

González de Frutos, U.: "Inteligencia artificial y Administración tributaria" en VV.AA.: *Fiscalidad e inteligencia artificial: Administración tributaria y contribuyentes en la era digital*, Serrano Antón, F. (Dir.), Ed. Aranzadi, Pamplona, 2020.

Gutierrez David, M.E., "Administraciones inteligentes y acceso al código fuente y los algoritmos públicos. Conjurando riesgos de las cajas negras decisionales", *Derecom*, 30, 2021.

Kanheman, D., Slovic, P. y Tversky, A.: "Judgment under Uncertainty: Heuristics and Biases", *Cambridge University Press*, 1982.

Llorente Sánchez-Arjona, M.: "Inteligencia artificial, valoración del riesgo y derecho al debido proceso" en VV.AA.: *Inteligencia artificial legal y administración de justicia*, Calaza López, S., Llorente Sánchez-Arjona, M. (Dirs.), Ed. Aranzadi, 2022.

López de Mántaras, R.: *Inteligencia Artificial*. Ed. Catarata, 2017.

Martín Diz, F.: "Modelos de aplicación de inteligencia artificial en justicia: asistencial o predictiva versus decisoria", *Justicia algorítmica y neuroderecho: Una mirada multidisciplinar*, Ed. Tirant lo Blanch, Valencia, 2021.

Martín López, J.: "Inteligencia artificial, sesgos y no discriminación en el ámbito de la inspección tributaria", *Artículos científicos, crónica tributaria*, núm. 182, 2022.

Mercader Uguina, J.R.: *Algoritmos e inteligencia artificial en el derecho digital del trabajo*, Ed. Tirant lo Blanch, 2022.

Monereo Pérez, J.L., Rodríguez Escanciano, S. y Rodríguez Iniesta, G.: "Algoritmos e inteligencia artificial. Implicaciones jurídicolaborales: Un enfoque desde la perspectiva de los derechos", *Revista Crítica de Relaciones de Trabajo*, Laborum. nº 8, 2023.

Muñoz Villareal, A., Gallego Corchero, V.: "Inteligencia artificial e irrupción de una nueva personalidad en nuestro ordenamiento jurídico ante la imputación de responsabilidad a los robots", *Inteligencia artificial y riesgos cibernéticos. Responsabilidades y aseguramiento*, Ed. Tirant lo Blanch, Valencia, 2019.

Nieva Fenoll, J.: *Inteligencia artificial y proceso judicial*, Ed. Marcial Pons, Madrid, 2018.

Oliver Cuello, R.: "Big Data e inteligencia artificial en la Administración Tributaria" en *Revista de Internet, Derecho y Política*, núm. 33, octubre 2021.

Peris Manguillot, A.: "Algoritmos ¿podemos hacerlos transparentes y trazables en su proceso? en VV.AA.: *Transparencia y explicabilidad de la inteligencia artificial*, Cotino Hueso, L., Castellanos Claramunt, J. (Eds.), Ed. Tirant lo Blanch, Valencia, 2022.

Planchadell-Gargallo, A.: "Inteligencia artificial y medidas cautelares" en VV.AA.: *Justicia algorítmica y neuroderecho. Una mirada multidisciplinar,* Barona Vilar, S. (Ed.), Tirant lo Blanch, 2022.

Valle Escolano, R.: "Transparencia en la inteligencia artificial y en el uso de algoritmos" en VV.AA.: *Transparencia y explicabilidad de la inteligencia artificial*, Cotino Hueso, L., Castellanos Claramunt, J. (Eds.), Ed. Tirant lo Blanch, Valencia, 2022.