

# INTELIGENCIA ARTIFICIAL Y VALORACIÓN DE LA PRUEBA: LAS GARANTÍAS JURÍDICO-CONSTITUCIONALES DEL ÓRGANO DE CONTROL

## ARTIFICIAL INTELLIGENCE AND THE ASSESSMENT OF EVIDENCE: THE LEGAL-CONSTITUTIONAL GUARANTEES OF THE SUPERVISING BODY

Pere Simón Castellano\*

Universidad Internacional de la Rioja (UNIR)

*Judicial proceedings and trials will be digital or they will not be. Overcoming the debate about the convenience of embracing the technological advance does not imply accepting integrated positions de facto. The COVID-19 pandemic has further highlighted the troubles and deficits of Spanish judicial proceedings. A slow justice, paralyzed due to remote work and without resources or guarantees to hold trials at a distance, is not justice. Excuses are running out today faster than ever, since new levels of effectiveness, efficiency, and speed, with full respect of procedural guarantees, can be reached with the use of most modern new technologies.*

*This article explores the possibilities of using artificial intelligence systems in a specific area: the assessment of evidence. In this context, the hypothesis that the article aims to demonstrate is that although the human presence will continue to be necessary in judicial proceedings, artificial intelligence systems could support the free assessment of evidence by the judge and cover those gaps where the human mind shows its most erratic side, such as during corroborations.*

**KEYWORDS:** *Artificial Intelligence; criminal proceedings; free assessment of evidence; Spain.*

*El proceso judicial será tecnológico o no será. Superar el debate acerca de la conveniencia de abrazar el avance tecnológico no implica aceptar de facto posiciones integradas. La pandemia del COVID-19 ha acentuado aún más –si cabe– las miserias y disonancias del proceso judicial en España. Una justicia lenta, paralizada ante la necesidad de teletrabajar y sin recursos con garantías para celebrar juicios a distancia, no es justicia, y hoy se agotan más rápido que nunca las excusas, puesto que nuevas cuotas de eficacia, eficiencia y celeridad, con pleno respeto a las garantías procesales, pueden alcanzarse con el empleo de las más modernas y nuevas tecnologías.*

*El presente trabajo pretende explorar las posibilidades de empleo de los sistemas de inteligencia artificial en un ámbito muy concreto: el de la valoración de la prueba. En este contexto, la hipótesis que el artículo pretende demostrar es que la presencia humana seguirá siendo necesaria, si bien los sistemas de inteligencia artificial podrían apoyar y asistir la libre valoración de la prueba por parte del juzgador y colmar aquellas lagunas donde el ser humano muestra su faceta más errática; por ejemplo, en las corroboraciones.*

**PALABRAS CLAVE:** *Inteligencia artificial; proceso penal; libre valoración de la prueba; España.*

\* Abogado. Doctor en Derecho por la Universitat de Girona. Exsocio fundador de Pridatect. Profesor contratado doctor en Derecho Constitucional en la Universidad Internacional de la Rioja (UNIR) (Logroño, España). Profesor de Derecho Penal Económico, Propiedad Industrial y Nuevas Tecnologías, Protección de Datos y Ciberdelincuencia en Universidad Internacional de La Rioja. Abogado independiente. Abogado en Font Advocats (Barcelona, España). Contacto: pere.simon@unir.net

Nota del Editor: El presente artículo fue recibido por el Consejo Ejecutivo de THĒMIS-Revista de Derecho el 27 de diciembre de 2020, y aceptado por el mismo el 8 de abril de 2021.

## I. INTRODUCCIÓN: DE LOS USOS POSIBLES DE LA INTELIGENCIA ARTIFICIAL EN LAS ADMINISTRACIONES PÚBLICAS Y EN EL SENO DEL PROCESO JUDICIAL

Internet lo ha revolucionado todo. La forma cómo nos comunicamos y relacionamos, la inmediatez en el acceso, la velocidad del mensaje y la importancia que le damos a la información que nuestras ‘amistades’ nos trasladan. En un mundo inundado de información irrelevante (Harari, 2018, p. 11), cada vez resulta más fácil descontextualizar lo que leemos y compartimos (Dumortier, 2009), subsumidos en una arquitectura en red, sin jerarquía, que pone en tela de juicio la aplicación de criterios y principios que tradicionalmente han permitido a los operadores jurídicos dar respuesta a aquellos casos difíciles en los que colisionan distintos bienes jurídico-constitucionales que bien merecen tutela (Simón Castellano, 2012; Cabellos Espiérrez, 2018).

En este complejo escenario de cambio constante al que nos aboca la postmodernidad líquida, surge con fuerza una nueva técnica que es capaz de predecir conductas humanas e, incluso, conocer mejor que el propio individuo cuáles son sus preferencias y necesidades, y, por ende, saber cuáles son las opciones o decisiones que más le conviene tomar. Nos referimos a los sistemas de inteligencia artificial (en adelante, IA), esto es, sistemas expertos que piensan y actúan como humanos, o bien piensan y actúan siguiendo la lógica racional. Sobre este particular, hemos analizado con mayor detenimiento los efectos de la IA en el seno de la administración de justicia en otro foro (Simón Castellano, 2021a).

Más allá de la precisión terminológica de lo que puede o debe entenderse por un sistema experto de inteligencia artificial, cuyo objeto de debate complicaría de sobremano no exceder la extensión de este trabajo, resulta necesario constatar, de un lado, (i) los cimientos en los que se basa la estructura de esta moderna técnica y, del otro, (ii) los campos y realidades tangibles en los que ya se puede apreciar un uso notable de algoritmos basados en el aprendizaje autónomo.

En relación con la primera de las categorías, el punto de partida para que podamos hablar de inteligencia artificial se encuentra en las técnicas que permiten el procesamiento masivo de datos, con independencia de si los podemos clasificar *stricto sensu* como datos personales. Sin tratamiento de datos ni sistema de procesamiento masivo, no existen sistemas inteligentes o, lo que es lo mismo, nadie puede proponer ni emular la actuación humana si opera en el vacío. Requiere, entonces, de

información y datos –cuantos más mejor– puesto que un mayor conocimiento y procesamiento permitirá obtener resultados más próximos a lo que el usuario realmente requiere o necesita, laminando y reduciendo así las posibilidades de que se produzcan sesgos por la poca calidad de la información o por la pobreza de los datos facilitados. El uso de algoritmos que disponen reglas lógicas para establecer una respuesta que emule, como decíamos anteriormente, la forma de pensar y actuar de los humanos (Mayer-Schönberger & Cukier, 2013, p. 13) necesariamente encuentra su *prius* formal y material en el procesamiento ingente de datos, personales o no.

Además, un matiz conceptual importante gira en torno a los llamados algoritmos avanzados que permiten el aprendizaje automático, más conocidos como sistemas de *machine learning*. Estos tienen un mayor potencial de impacto y una mayor fiabilidad, por la capacidad de aprender de los datos y de la experiencia. La principal arista del empleo por parte de los poderes públicos de este tipo de sistemas de aprendizaje profundo es la falta de transparencia (Cerrillo i Martínez, 2019) en un algoritmo que muta o se transforma, que se modifica y perfecciona, de forma autónoma, sin que el usuario sea capaz de descubrir de forma sencilla las razones por las que este ha adoptado o propuesto una decisión o resultado.

Por lo que se refiere a los usos tangibles de la inteligencia artificial, cabe señalar que, actualmente, los sistemas de inteligencia artificial son empleados en los ámbitos del procesamiento del lenguaje natural, la visión computarizada y la robótica, para fines muy diversos (Kaplan, 2017, p. 53). Los algoritmos están muy presentes en nuestra realidad cotidiana. Especialmente han sido empleados con notable fortuna, aunque también con una fiabilidad a vueltas bastante pobre en procesos de *marketing online* y publicidad comportamental, y condicionan con creces las decisiones que se toman, por ejemplo, en el ámbito del consumo de productos y servicios en línea. Para algunos autores, somos ya esclavos de la tiranía de los algoritmos (Edwards & Veale, 2017, pp. 19-20).

Una tiranía, en cierto modo, cada vez más tangible, puesto que el empleo de la inteligencia artificial está en claro auge (Cerrillo i Martínez & Peguera Poch, 2020) o, como algunos autores afirman, “está de moda” (De Asís Roig, 2018, p. 28). Un corolario que no resulta difícil de alcanzar si tenemos en cuenta la atención creciente que los mercados le están prestando, así como con los informes de prestigiosas firmas que le auguran un crecimiento imparable a lo largo de la próxima década (Eggers *et al.*, 2018, p. 3). Se manifiesta que la IA trans-

formará definitivamente la forma en la que vivimos, y proyectará sus efectos sobre las relaciones del trabajo y la producción automatizando tareas manuales o rutinarias, en el marco de una nueva, llamada cuarta, revolución industrial que vendrá acompañada de un nuevo paradigma en el que el aprendizaje constante y la especialización no serán una opción (Frey & Osborne, 2017, p. 38).

Sin embargo, más allá de su crecimiento imparable, ¿qué fines plausibles pueden tener los sistemas de inteligencia artificial en las administraciones públicas, en general, y en la administración de justicia, en particular? ¿Qué transformaciones puede proyectar sobre las estructuras de instituciones tan sensibles como los principios y garantías del proceso judicial?

Ya disponemos, en la actualidad, de numerosos ejemplos del empleo de sistemas autónomos de inteligencia artificial en las administraciones públicas: Atlanta predice el riesgo de incendio de los edificios; Hampton, el riesgo de riadas; y Chicago y Las Vegas identifican los locales que serán objeto de inspección, entre muchos otros (De Souza, 2018). Otras administraciones han empezado a utilizar esta moderna técnica para detectar fraudes a la seguridad social; para alertar del conflicto de intereses y evitar la corrupción en los procesos de licitación y contratación pública; con el fin de asistir a médicos para el tratamiento de enfermedades; para asignar subvenciones o evaluar profesores; para detectar el abandono escolar, estableciendo alertas automáticas basadas en el comportamiento de empleados públicos, etc.

Como reacción, la doctrina ha estudiado con profundidad tal posibilidad (Martínez Martínez, 2019; Valero Torrijos, 2019; Cotino Hueso, 2019; Boix Palop, 2020), si bien con carácter general para todas las administraciones públicas y desde el prisma de los efectos y retos jurídicos que plantea la inteligencia artificial.

El legislador, por su parte, a vueltas se ha posicionado abrazando tal posibilidad, incluso con demasiada premura y sin suficientes cautelas. Es el caso de la Ley 22/2018, de 6 de noviembre de 2018, de la Generalitat València, que prevé un sistema de alertas tempranas anticorrupción (en adelante, SATAN) basado en los criterios de Falciani, respecto del que no tenemos constancia alguna de que se esté utilizando. La Agencia Española de Protección de Datos (en adelante, AEPD), en el Informe

385661/2017 (2018), y la doctrina (Cotino Hueso, 2020) han valorado de forma crítica la citada ley, pues consideran que restringe de forma inadmisibles los derechos al no cumplir los requisitos constitucionales respecto a sus límites ni los establecidos en el artículo 23 del Reglamento General de Protección de Datos de la Unión Europea (en adelante, RGPD). La celeridad de la respuesta del legislador no siempre va acompañada de la necesaria calidad, y, por ello, también disponemos de la primera resolución judicial que ha declarado ilegal un algoritmo sobre evaluación de características personales de los ciudadanos. Nos referimos a la Sentencia de 5 de febrero de 2020 del Tribunal de Distrito de la Haya, asunto C/09/550982/HA ZA 18-388 (*NJCM et al./Países Bajos*, 2020; Fernández Hernández, 2020c; Cotino Hueso, 2020).

Frente a la asimetría y teniendo en cuenta las primeras experiencias contradictorias, el legislador europeo parece decidido a aportar cierta seguridad jurídica al respecto. El 20 de octubre de 2020, el Pleno del Parlamento Europeo aprobó su propuesta de regulación de la inteligencia artificial, cuya principal finalidad es impulsar la innovación a partir de normas éticas que promuevan la confianza en la tecnología<sup>1</sup>.

La propuesta, que nace del Parlamento Europeo y tiene como destinatario a la Comisión Europea, se refiere a tres áreas básicas: las normas éticas que deben regir la inteligencia artificial, las reglas sobre la responsabilidad por las consecuencias derivadas de su uso y, finalmente, las reglas sobre derechos de propiedad intelectual vinculadas al uso de los sistemas inteligentes. Tal regulación, esperada en forma de reglamento, persigue garantizar cierta uniformidad de las futuras normas que se elaboren sobre esta materia en el marco de la Unión Europea, al establecer principios rectores que exigen que la IA se desarrolle desde un enfoque antropocéntrico, es decir, centrada y hecha por el ser humano. También, se busca que se garantice la seguridad, transparencia y responsabilidad derivada de su uso, así como la disponibilidad de salvaguardias contra la parcialidad y la hipotética o eventual discriminación del algoritmo. Asimismo, que incorpore el derecho a la reparación en caso de que se produzcan daños ilegítimos y exige la responsabilidad social, ambiental y el respeto de la privacidad y la protección de datos. Sea como fuere, se trata de un proyecto en tramitación y habrá que estar atentos a su recepción por parte de la Comisión Europea y su futura aprobación.

<sup>1</sup> Véase la Resolución del Parlamento Europeo, 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 (2020/2012(INL)).

Por lo que aquí interesa, y retomando el camino trazado, corresponde analizar los usos de la IA en el seno del proceso judicial y en la administración de justicia, lo que puede tener mucha lógica y sentido, por ejemplo, para la propia gestión del seguimiento de archivos, con indiferencia de la naturaleza del proceso o del orden jurisdiccional (Simón Castellano, 2021a). La creación de alertas tempranas por parte del algoritmo en función de la inmensa cantidad de datos procesados y la experiencia previa que este haya acumulado puede contribuir a simplificar su gestión y aumentar la eficiencia (European Commission For The Efficiency Of Justice [CEPEJ], 2018, p. 63). Además, los discursos en torno a la digitalización y modernización de la justicia mediante sistemas de IA van mucho más allá de la eliminación del papel, cuya importancia radica en que la información ahora está digitalizada y es 'tratable', con un potencial enorme para fines muy diversos, desde la estadística hasta la gestión, aunque todo redunde en la agilización de la justicia y la automatización de buena parte de la labor judicial (Ercilla García, 2020).

Siguiendo en el plano general, las técnicas de IA pueden mejorar el acceso a la jurisprudencia sobre la base del aprendizaje automático, y son un activo considerable para complementar búsquedas o vincular varias fuentes legales y jurisprudenciales, incluso mediante técnicas de visualización de datos (CEPEJ, 2018, p. 63). También la inteligencia artificial puede contribuir a fortalecer la relación entre la justicia y los ciudadanos, haciendo a la primera más transparente gracias a los *chatbots*, que pueden configurarse para facilitar el acceso a las diversas fuentes de información existentes utilizando lenguaje natural, compartiendo información (plantillas de documentos, solicitudes o modelos contractuales básicos) o incluso generándola en línea (CEPEJ, 2018, p. 63).

La inteligencia artificial, en la jurisdicción civil y social, debería tener mucho recorrido en la llamada resolución extrajudicial de conflictos y controversias, ámbito en el que el empleo de sistemas expertos solo puede redundar en beneficios para todas las partes, así como en el ámbito del llamado *Online Dispute Resolution* (en adelante, ODR). Sirva como ejemplo el uso de técnicas de justicia predictiva por parte de las compañías aseguradoras, que calculan y evalúan las posibilidades de éxito de acudir a los tribunales y, en el caso que este resulte inferior a un determinado porcentaje, prefieren acudir a sistemas de resolución extrajudicial de conflictos, evitando los costes del proceso judicial.

Otros usos, inclusive más cuestionables, de la inteligencia artificial son los relativos al perfilado de jueces y magistrados, y los destinados a anticipar,

en sentido literal, el sentido de las resoluciones judiciales. Cuantificar y analizar exhaustivamente la actividad de un juez contribuiría a disponer de unos resultados que, en el mejor de los casos, podría considerarse un asesoramiento cuantitativo y cualitativo para los propios jueces, con fines puramente informativos y de asistencia en la toma de decisiones (CEPEJ, 2018, p. 66). Así, el análisis por parte de sistemas expertos de la predictibilidad y previsibilidad de las decisiones judiciales estaría orientado exclusivamente a reducir el margen de discrecionalidad de los jueces, considerado en ocasiones excesivo (Fernández Hernández, 2020b). Estas herramientas, empero, podrían estar abiertas al abuso si se emplean no solo para gestionar los asuntos judiciales de manera más eficiente, sino también para evaluar el grado de desempeño de los jueces, incluido el análisis de supuestos sesgos ideológicos en sus patrones de comportamiento.

Los esfuerzos en el ámbito de la jurisdicción penal se pueden englobar, básicamente, en los fines relacionados con la actividad predictiva —o *predictive policing*— y en el llamado 'reconocimiento facial' (González Fuster, 2020). En cuanto a la actividad predictiva, destacamos la posibilidad de apoyarnos en el avance tecnológico con el fin de revelar patrones de una probable comisión de actos delictivos o a la identificación de posibles víctimas de futuros delitos, permitiendo su intercepción antes de que ocurran al más puro estilo que mostró al mundo el paradigma cinematográfico de *Minority Report*. La predicción del crimen es una tendencia imparable cuya *vis* expansiva ha marcado notablemente los avances en materia de seguridad a nivel mundial y en Europa desde los atentados del 11 septiembre de 2001 (Souza de Menezes & Agustina, 2020), con todos los problemas asociados a la calidad de la predicción, que necesariamente debe estar basada en una base de datos suficiente, en términos cuantitativos y cualitativos (Richardson *et al.*, 2019).

El uso de estos instrumentos de predicción por parte de los jueces de la jurisdicción penal en Europa es muy escaso (González Fuster, 2020); no obstante, ya disponemos de resultados de experiencias significativas en modelos en perspectiva comparada, más concretamente, en Estados Unidos y Reino Unido. Me refiero a los programas *Correctional Offender Management Profiling for Alternative Sanctions* (en adelante, COMPAS) y el *Harm Assessment Risk Tool* (en adelante, HART), que calculan el riesgo de reincidencia de una persona que ha sido anteriormente condenada.

Lo cierto es que las principales conclusiones o propuestas formuladas por los citados algoritmos mostraron un preocupante sesgo; por ejemplo, en

el caso de COMPAS, se demostró la existencia de un sesgo racial indirecto en los modelos que predicen el riesgo de reincidencia mediante el uso de variables sustitutivas que no son neutrales. Aunque no es pacífico, estudios afirman que, aplicando el citado algoritmo, los acusados negros tenían casi el doble de probabilidades, en relación con los blancos, de ser considerados en situación de alto riesgo de reiteración delictiva (Babuta *et al.*, 2018, p. 7). Los proyectos COMPAS y HART han sido objeto de duras críticas por las organizaciones no gubernamentales ProPublica y Big Brother Watch, las que subrayan los problemas que se proyectan por aplicar un enfoque discriminatorio o determinista en el diseño del algoritmo.

En cuanto a las tecnologías de reconocimiento facial, estas son objeto de notable controversia y limitaciones por lo que se refiere a su utilización, especialmente por parte de empresas privadas. Lo cierto es que se han implementado, hasta la fecha, muchas iniciativas de reconocimiento facial en países como China, Rusia, Estados Unidos, Suecia, entre otros, tanto por parte de autoridades como por parte de empresas privadas (González Fuster, 2020, pp. 24-25). La idea es prevenir el delito o dar cumplimiento a prohibiciones emitidas por sujetos de derecho público o privado, mediante la identificación con reconocimiento facial del infractor. Ciertas iniciativas han obtenido un beneplácito expreso por parte de las autoridades nacionales de protección de datos, como es el caso de la autoridad de protección de datos danesa, que autorizó el empleo de sistemas de reconocimiento facial por parte de un club de fútbol para evitar la entrada en el estadio de personas que tenían el acceso prohibido (IT-Pol, 2019).

Por su parte, los sistemas de Inteligencia Artificial Judicial de Valoración de Riesgo (en adelante, IAJVR), siguiendo la terminología propuesta por Miró Llinares (2018, p. 107), son otra opción con una funcionalidad muy distinta a las anteriores, la mayoría apoyados hoy en día en sistemas expertos *Human in Command*, no basados en una *black box* o en procesos de aprendizaje autónomo o *deep learning*. Los sistemas que emplean la IAJVR están basados en la automatización de las herramientas actuariales de valoración del riesgo y son aplicables a los procesos de toma de decisiones que requieren conjurar el riesgo, por ejemplo, en la aplicación de medidas cautelares o, tras la condena, en el ámbito penitenciario y en la ejecución penal.

Se trata de un concepto clave que hemos tenido la ocasión de estudiar en profundidad, por lo que se refiere a potenciales usos, ventajas e inconvenientes, así como límites, en relación con la medida cautelar más polémica, la prisión provisional, en otro foro (Simón Castellano, 2021b).

En resumen, y en términos generales, la inteligencia artificial se utiliza o se está estudiando para utilizarla en los sistemas jurídicos europeos con diversos fines, entre los que destacan los de facilitar el acceso a la justicia (por ejemplo, mediante los *chatbots*), apoyar medidas alternativas de solución de conflictos en el ámbito civil, para el perfilado de jueces (Council Of Bars And Law Societies Of Europe [CCBE], 2020), para el análisis predictivo y prevención *ex ante* del delito y, finalmente, para el reconocimiento facial del infractor.

## II. INTELIGENCIA ARTIFICIAL, VALORACIÓN DE LA PRUEBA Y ARGUMENTACIÓN PROBATORIA

Hemos analizado hasta este punto los usos y finalidades plausibles de los sistemas de inteligencia artificial, tanto en las administraciones públicas, en general, como en la administración de justicia, en particular. Intencionadamente, no mencionamos hasta el momento a la posibilidad de que los sistemas expertos puedan ser empleados *ad hoc* de la valoración de la prueba y la argumentación probatoria en sede judicial.

Constituye al respecto una obra de referencia obligada, el trabajo Inteligencia artificial y proceso judicial del Dr. Nieva Fenoll (2018), que anticipa de una forma muy acertada el impacto de los sistemas de IA en la valoración de la prueba. Lejos de establecer un sistema de automatismos nada respetuoso con la vigencia de nuestro modelo de libre valoración de prueba<sup>2</sup>, por contraposición a otros modelos en perspectiva comparada como el de las *hearsay rules* de Estados Unidos, la IA podría, en cambio, contribuir a ofrecer resultados basados en datos bastante objetivables en función de los medios de prueba empleados (Nieva Fenoll, 2018). Todo ello, como decíamos, también en sistemas de libre valoración de ella, lo que ya de entrada implica una apreciación probatoria más compleja.

Todo proceso penal estructurado con base en una concepción racionalista de la prueba debe tener como objetivo darle al juez las herramientas

<sup>2</sup> Cuando hablamos de la libre valoración de la prueba nos referimos a la regla que permite al órgano judicial competente apreciar las pruebas existentes en el procedimiento según su libre convencimiento y sin tener que otorgarles a algunas de ellas un valor o credibilidad superior que venga predeterminada por la ley (Diccionario panhispánico del español jurídico, 2020). Esta es la definición que ofrece el Diccionario panhispánico del español jurídico en el que colaboran la Real Academia Española, la Cumbre Judicial Iberoamericana y la Asociación de Academias de la Lengua Española.

necesarias para que la valoración de la prueba dependa, en la medida de lo posible, de criterios objetivos y explicables (Ferrer Beltrán, 2002; Ferrer Beltrán, 2007), y no de instituciones generales cubiertas detrás del tantas veces incoherente ‘telón’ de la intermediación (Taruffo, 1992/2002; Nieva Fenoll, 2010). Es en este ámbito, y para este caso concreto particular, en el que debemos valorar la posibilidad real de que los sistemas de IA ofrezcan una oportunidad, también, para expulsar de las salas de justicia los atávicos heurísticos judiciales o, lo que es lo mismo, las emociones, gustos, prejuicios, preferencias, aficiones y otros elementos que no siempre resultan confesables y que están muy presentes en la valoración que realiza el juzgador.

#### A. ¿Cómo valoran las pruebas los jueces?

Solo si comprendemos como valoran las pruebas los jueces seremos capaces de articular una fórmula algorítmica que emule o replique su proceso lógico y psicológico de valoración probatoria. Los seres humanos somos automáticos, recurrentes, en una mayor medida de lo que podemos pensar, en nuestro día a día, en nuestro actuar cotidiano. Como señalan Kahneman y Sunstein, nos gusta repetir nuestras experiencias y ritos más ancestrales (en HarvardCPL, 2014). Los jueces también son personas y no podrían escapar a algo tan humano como la forma en la que se accede al conocimiento y cómo este, sumado a otros factores, influye decisivamente en las decisiones que luego tomamos. Los humanos somos automáticos y los jueces, por ende, también son mecánicos en la mayoría de las ocasiones.

Por si lo anterior fuera poco, además, los jueces no operan en el vacío. Ni pueden, ni deben hacerlo. Lo que sucede en el exterior nos condiciona, nos aflige y llega a afectar nuestra toma de decisiones. Es lo que Kahneman (en HarvardCPL, 2014) llama el efecto del contexto y enlaza con la aficción y la emoción. O lo que fue, en su día, brillantemente expuesto por Chayes (1976), cuando nos advertía que el juzgador no puede actuar aislado del entorno social y las decisiones que este toma en los litigios de interés público son fruto de un diálogo natural, permanente e inevitable con otros elementos sociales y políticos –académicos, prensa, el Poder Ejecutivo, el legislador, agencias administrativas y el público en general–.

Los jueces también tienen ideología y se rigen por motivos ideológicos (Posner, 2011). Sería falaz no reconocerlo, aunque otra cosa bien distinta es que hayan recibido una formación que les permite tomar decisiones basadas en la ley y en la interpretación que un jurista puede hacer de esta. Esta

formación profesional de jueces y magistrados actúa como límite interno a los atajos intuitivos más salvajes de nuestra psicología cognitiva. Señala Nieva Fenoll (2018, p. 44) que la carga de trabajo o la sobrecarga, mejor dicho, acostumbra a provocar un alejamiento de las realidades a juzgar y empuja al juzgador a dejarse llevar por su ideología.

Así las cosas, los jueces toman decisiones en el marco de un ejercicio cognitivo que vincula la experiencia –vital y profesional–, las emociones y el contexto, el diálogo con otros elementos políticos y sociales, la ideología base, los heurísticos de pensamiento descritos por Kahneman, Slovic y Tversky (1982), la memoria y estadística *prima facie* –es decir, aquello que recuerdan y que les viene a la cabeza de forma casi automática sobre casos anteriores–, la intuición (Myers, 2002, p. 16), la fatiga –mental y física– y la comodidad–placer, satisfacción e incluso el éxito pretérito en situaciones análogos–.

¿Podemos trasladar al algoritmo este proceso de valoración humana del juzgador, de los hechos, y hacerlo, además, expulsando o mitigando los atávicos heurísticos judiciales? En otro foro, y en relación con la adopción de la siempre polémica medida de prisión provisional, hemos defendido enérgicamente una respuesta afirmativa (Simón Castellano, 2021b). Veamos, empero, cómo podría extrapolarse o traducirse todo ello al ámbito de la valoración probatoria.

#### B. La valoración de la declaración de las personas

Las leyes procedimentales incorporan diferentes pruebas –testifical, declaración de las partes, etc.– que podrían englobarse dentro de una categoría superior, de declaración de las personas. ¿Qué podrían aportar los sistemas de IA en este contexto? Hoy en día, los jueces valoran la credibilidad de las personas de acuerdo con los factores situacionales y en función de su experiencia vital anterior. Sin embargo, la memoria humana es finita y limitada, así que carece de precisión matemática y los cálculos estadísticos humanos, fundamentados en metaanálisis, suelen acabar incurriendo en sesgos de confirmación, de comodidad y anclaje o de discriminación, entre otros. Las circunstancias que afectan a la credibilidad del testimonio de las personas son parcialmente objetivables (Nieva Fenoll, 2018, p. 81), lo que significa que estas podrían digitalizarse o traducirse en lenguaje de programación.

El problema pasa por el cómo, puesto que construir una herramienta que recoja todos los datos del caso concreto y que los combine adecuadamente para valorar la credibilidad de la declara-

ción no es una tarea fácil. Buena muestra de ello es el programa ADVOKATE, empleado para valorar la credibilidad de los testigos, que se basa en la distancia del observador, la visibilidad, el conocimiento previo entre el testigo y la persona observada, la competencia y conocimiento que se atribuye al testigo –edad, estudios, capacidad mental, etc.–, si se detectan errores en la descripción y contraste de la realidad, el tiempo de observación y el tiempo discurrido entre los hechos y la declaración (Bromby & Hall, 2002, pp. 143-152).

Como se observa, se trata de fórmulas que pueden proyectarse sobre un algoritmo, sobre bases estadísticas y sistemas de cálculo que objetivan la intuición humana y la perfeccionan, acercándonos más a la realidad de los hechos. Estas fórmulas deberían incluir, siguiendo a Nieva Fenoll (2018, p. 84 & ss) (i) la coherencia de la declaración; (ii) la contextualización o capacidad del interrogado de recuperar la descripción del escenario en el que sucedieron los hechos; (iii) la corroboración en otros medios de prueba de lo declarado; y (iv) la existencia o no de comentarios extra, oportunistas, que solo buscan reforzar retóricamente la declaración y que, a la postre, nos indican que el declarante puede tener algún fin metajurídico o voluntad de proyectar una realidad determinada al juzgador. En el proceso habría que descartar expresamente el estudio gestual (Ekman, 2011). De todos ellos, el parámetro de la corroboración es el que con mayor facilidad puede ser evaluado a través de la IA, puesto que el ser humano es incapaz de procesar a gran velocidad y chequear o comprobar la coincidencia o inexactitud de lo que se declara con las otras pruebas practicadas. Así las cosas, la IA sería de gran utilidad en las corroboraciones, en las que el ser humano se muestra más débil por las dificultades inherentes de su procesamiento.

Nieva Fenoll (2018, pp. 88-89) señala también como fin potencial de la IA el que esta pueda ser aplicada para descartar preguntas dirigidas por parte de los interrogadores o preguntas que de fondo impliquen valoraciones profundamente subjetivas, que obligan de facto al declarante a asumir inconscientemente ciertas premisas, excluyendo así las preguntas no neutrales del interrogatorio. Finalmente, la aplicación de la IA a la neurociencia en el seno del proceso penal (Balbuena Pérez, 2017; Julià Pijoan, 2020) es otro campo de gran recorrido, si bien este ámbito debería ser explorado con las mayores cautelas posibles por la imposibilidad de detectar los falsos

recuerdos y las mentiras inconscientes que son, precisamente, la mayoría de las que se escuchan en un proceso judicial.

### C. La valoración de la documental aportada

Resultan indubitados los efectos benéficos del empleo de sistemas de IA para la valoración de los documentos, sea cual sea el soporte y el formato en el que se presente la información. Ténganse en cuenta los sistemas de procesamiento masivo de datos, o *big data*, y las posibilidades de analizar ingentes cantidades de información de forma lógica y ofrecer los resultados de forma estructurada. No se trata solo de capacidad de sumar grandes cantidades de información, sino también de la posibilidad de analizar con eficacia el lenguaje del documento, el vocabulario, las expresiones, la fiabilidad y autenticidad, así como técnicas que avanzan en la comprensión de un documento emulando el proceso cognitivo humano. Los sistemas de IA pueden comprender el significado de los textos, sea cual sea su idioma, y procesarlo ofreciendo una visión que ayudará a complementar la interpretación y valoración humana del juzgador, que ahora puede apoyarse en estas modernas técnicas.

### D. La valoración de la prueba pericial

También, en el ámbito de la prueba pericial muchas son las posibilidades en relación con el empleo de los sistemas expertos de la IA. La pericial permite al juzgador, como jurista, acercarse a un ámbito del conocimiento o a un ámbito científico que desconoce profundamente. Lo hará sobre la base de la pericia y conocimiento de un experto de ese sector, apoyándose en sus resultados de forma crítica, aplicando distintos criterios de valoración, tanto del currículo del perito, como sustantivos, de fondo, como los llamados criterios Daubert de admisión de prueba pericial (Vázquez Rojas, 2015). Unos criterios que se construyen jurisprudencialmente en tres sentencias conocidas como la trilogía Daubert<sup>3</sup> y que han supuesto la modificación, en 2011, del artículo 702 de las *Federal Rules of Evidence*, cuyos efectos se proyectan incluso en los criterios doctrinales y jurisprudenciales que ayudan al juez español a valorar la prueba pericial más allá de la regulación sobre la valoración de la prueba contenida en el artículo 348 de la Ley de Enjuiciamiento Civil (2000) (en adelante, LEC).

Respecto a la valoración del currículo, es evidente que los sistemas de procesamiento masivo pueden

<sup>3</sup> Nos referimos a las sentencias de la Corte Suprema de los Estados Unidos, *Daubert c. Merrell Dow Pharmaceuticals* (1993); *General Electric Co. c. Joiner* (1997); *Kumho Tire Co. c. Carmichael* (1999).

indicar información muy relevante, detectando plagios o descubriendo méritos aparentes. Más allá de la creación de fórmulas o reglas lógicas que detallen un porcentaje de *expertise* determinado, los sistemas de IA también pueden leer los trabajos y aportaciones previas del supuesto experto para determinar si este realmente lo es, identificando la creatividad y originalidad de sus aportaciones (Nieva Fenoll, 2018, pp. 94-95).

En relación con los criterios Daubert, el principal reto es el digitalizarlos o automatizarlos. Parece evidente que un sistema experto de IA, en su modalidad más simple, de aplicación de algoritmos con reglas lógicas, bien podría resultar de utilidad a tal efecto, con los efectos benéficos de poder explicar el funcionamiento de la máquina con transparencia y publicidad. Por el contrario, un modelo de aprendizaje autónomo profundo o *deep learning* no permitiría tal extremo, al evolucionar el algoritmo dentro de la llamada *black box*. De esa manera, una vez expulsada la posibilidad de los sistemas de *deep learning*, nos queda observar el núcleo de lo que se pretende automatizar. Los criterios Daubert podrían simplificarse en el estudio o análisis de cinco ítems en relación con la prueba pericial: que la técnica siga un método científico verificado empíricamente; que sea pública y haya sido objeto de revisión por otros expertos; que indique el grado o porcentaje de error; que mantenga estándares y controles sobre la fiabilidad de la técnica; y que exista cierto consenso en la comunidad científica acerca de la técnica empleada. Como se observa, son criterios que podrían ser aplicados por sistemas expertos de IA programados *ad hoc*; de hecho, como señala Nieva Fenoll (2018, p. 97), parecen haber sido configurados realmente para construir un algoritmo sobre ellos.

### III. TRANSPARENCIA DEL ALGORITMO E INDEPENDENCIA JUDICIAL: PAUTAS Y CRITERIOS PARA LA CREACIÓN DE UN ORGANISMO DE CONTROL

El empleo de sistemas expertos de IA en sede judicial nunca sustituirá a la persona (Nieva Fenoll, 2018; Simón Castellano, 2021b). Principalmente, porque no pueden superar el mandato constitucional del artículo 117 de la Carta Magna española (Constitución española, 1978), el cual señala que solo jueces y magistrados pueden juzgar y hacer ejecutar lo juzgado. Un androide o máquina no puede —por ahora—, ni debe —en el futuro—, hacerlo. Sin embargo, lo anterior no significa que no se puedan emplear sistemas expertos de IA como apoyo a determinados procesos judiciales, como para la adopción de decisiones tan significativas como la prisión provisional (Simón Castellano, 2021b), la libertad condicional o para la valoración

de la prueba. Mientras la función jurisdiccional esté reservada a jueces, siempre necesitaremos contar con la participación decisiva del juzgador, quien, además, debe ser el humano al frente del sistema experto de IA en el ámbito de los llamados sistemas *Human in Command* (en adelante, HIC) o, en caso contrario, debe poder decidir si aplica y cómo los resultados y proyecciones que este ofrezca.

En este ámbito, los sistemas de IA no pueden ser ni implicar una *black box*, tampoco serían admisibles los sistemas técnicos basados en aprendizajes autónomos o *deep learning*. Estos avances no son aplicables a la institución procesal por tratarse de técnicas que no permiten acceder con publicidad y transparencia real al código, que cambia o muta constantemente en función de diferentes parámetros y que, además, no es explicable ni necesariamente razonable desde la óptica humana. No se trata tanto de que esos códigos no puedan publicarse, sino que, si se trata de sistemas de aprendizaje autónomo, probablemente no vamos a ser capaces de explicar o comprender las reglas lógicas, si es que estas existen, que hay detrás del mismo. Ni siquiera su creador o programador, es decir, quién lo diseñó en origen y quién escribió el código, sería capaz de explicar los resultados, valoraciones y decisiones que toma la máquina.

#### A. Sesgos de la IA y problemas de aplicación práctica

Veamos un ejemplo de cómo los sistemas de IA se han introducido en el ámbito del derecho penal sustantivo y procesal para determinar cuestiones tan relevantes como para fundamentar la libertad condicional, la condena a una pena de prisión de larga duración o la imposición de una medida de seguridad (Romeo Casabona, 2018, pp. 48-55.; Martínez Garay, 2018). Me refiero al caso *Estado de Wisconsin c. Loomis* (2016), en el que el Tribunal Supremo de ese Estado resuelve el caso de un particular, Eric Loomis, que había sido acusado de cinco delitos por participación en un tiroteo efectuado desde un vehículo. Para lo que aquí interesa, el juez de instancia aceptó la conformidad del investigado y al concretar la pena tuvo en cuenta, entre otras cosas, el contenido de un *presentence investigation report* basado en el resultado de la evaluación de Eric Loomis mediante la herramienta COMPAS, que, como se mencionó anteriormente, utiliza IA. Esta arrojaba con relación a Eric Loomis un riesgo alto de reincidencia, en general, y de reincidencia violenta, en particular. El juez condenó a Loomis a una pena de seis años de prisión y otros cinco de supervisión post penitenciaria, razonando que la pena es el resultado de la valoración de distintos factores, entre los que se encuentra el riesgo extremadamente alto de reincidir que le



atribuye la herramienta COMPAS, excluyendo la posibilidad de suspender la condena<sup>4</sup>.

Como reacción, Eric Loomis, a través de su representante, interpuso recurso ante el Tribunal Supremo del Estado de Wisconsin, alegando que el uso de la herramienta de IAJVR había vulnerado su derecho a un juicio justo por basarse en información poco fiable o precisa por tratarse de una decisión automatizada, no individualizada a su caso concreto, y por haberse tenido en cuenta indebidamente su género para determinar la pena (De Miguel Beriaín, 2018). Ninguno de los motivos del recurso prosperó. El Tribunal Supremo de Wisconsin aceptó la utilización de los sistemas de IAJVR y su adecuación a las garantías constitucionales argumentando que el derecho de los acusados a un juicio justo no fue violado por el mero hecho de que la defensa no pudiera acceder a una explicación adecuada sobre el tratamiento o la fórmula de los algoritmos del sistema COMPAS (*Estado de Wisconsin c. Loomis*, 2016). Para apoyar esta posición, el Tribunal señaló que la exactitud de los instrumentos utilizados y la capacidad de los jueces para entender su posible mal funcionamiento eran suficientes para asegurar los derechos de los acusados (*Estado de Wisconsin c. Loomis*, 2016, párr. 54-57; 68-72; 83-86). De hecho, indica el Tribunal Supremo de Wisconsin que los sistemas de IAJVR aportan al juzgador información que, unida al resto de circunstancias concurrentes, contribuyen a que la condena se ajuste lo más posible al caso concreto. Lo que los jueces deben hacer cuando manejen estas herramientas es aplicar a sus resultados un margen de discrecionalidad, poniéndolos en el contexto del resto de circunstancias particulares propias del caso y del individuo concreto (Martínez Garay, 2018, p. 492).

Una 'solución' como la señalada no sería viable en un modelo más tuitivo, garantista, desde la óptica del derecho de defensa, como es el español. Los efectos de la sentencia *Estado de Wisconsin c. Loomis* son realmente preocupantes, puesto que esta práctica ha sido aceptada y confirmada en otros Estados. Además, como bien señala el voto particular a la sentencia, esta doctrina es muy insatisfactoria al producirse en un contexto en el que cada vez más las decisiones serán materialmente tomadas, parcial o totalmente, por los algoritmos y la programación basada en el código fuente aprobado (Romeo Casabona, 2018, p. 39-55). Sea como fuere, la referencia nos sirve como muestra de los problemas que proyecta el empleo de sistemas de

IA en el seno del proceso penal. ¿Cómo se defiende el acusado o el investigado de una decisión que se apoya en un algoritmo opaco, cuyo funcionamiento no es transparente y cuyas reglas lógicas son desconocidas? ¿Qué sucede si los sesgos humanos se han proyectado o visto incluso reforzados por el código de la máquina? A nuestro modo de ver, y así lo hemos defendido a profundidad (Simón Castellano, 2021b), la respuesta pasa por la definición de un órgano de control público que garantice los principios de publicidad, transparencia, trazabilidad, auditabilidad y explicabilidad del código, y, muy significativamente, el principio de jurisdiccionalidad y legalidad desde el diseño y por defecto.

#### **B. El órgano de control como garante de la independencia judicial y los principios de naturaleza jurídico-constitucional**

El papel del órgano de control es fundamental desde el nacimiento hasta el funcionamiento y gestión de los sistemas expertos de IA que, eventualmente, puedan habilitarse para la valoración de la prueba en sede judicial. Recientemente, en el Boletín Oficial del Estado (en adelante, BOE) de 10 de diciembre de 2020 se publicó la Orden JUS/1179/2020, de 4 de diciembre, por la que se modifica la Orden JUS/1362/2016, de 3 de agosto, por la que se crea la Comisión Ministerial de Administración Digital del Ministerio de Justicia (en adelante, la Comisión) y se regula su composición y funciones. Un órgano que, a nivel nacional, parece estar predestinado a ejercer el control sobre los sistemas de IA en los que se puedan apoyar jueces y magistrados. Se quiere digitalizar la administración de justicia y mejorar la eficacia y eficiencia de los servicios, y el principal problema que se ve es que la lista o composición del órgano sorprende por sus ausencias notables: ningún juez, ni fiscal, ni letrado de la administración de justicia, ni abogado, ni funcionario. Al margen del representante de la Abogacía General del Estado, la Comisión está integrada por personas legas, que no saben ni conocen qué es un procedimiento judicial, ni lo han visto, ni saben los problemas diarios de los juzgados<sup>5</sup>. Tomarán decisiones e impulsarán desde el más absoluto vacío, el de la ignorancia y falta de conocimientos sobre qué necesitan nuestros juzgados y tribunales. No se trata de corporativismo ni de nada parecido. Si queremos introducir modernas tecnologías en sede judicial parece evidente que deberíamos recabar y tener muy presentes en la Comisión que impulsa esa transformación digital a los principales actores del proceso judicial.

<sup>4</sup> Véase la Sentencia de *Estado de Wisconsin c. Loomis* (2016, párr. 19 & ss).

<sup>5</sup> En esta misma dirección lo ha criticado la Asociación Profesional de la Magistratura (2020) en su cuenta de Twitter.

Precisamente, el principio de jurisdiccionalidad y legalidad desde el diseño del sistema de IA es el que exige que jueces y magistrados, los principales actores de la valoración probatoria, participen de forma activa en el órgano de gestión y control de aquel. De esa manera, cuando hablamos de un órgano de control de los sistemas de IA en el ámbito judicial, el órgano de gestión pasa, a *contrario sensu* de la realidad práctica de rabiosa actualidad, necesariamente, por la implicación y participación de todos los actores y agentes implicados en el proceso judicial. Una participación que debe producirse de forma integral cuando se trata de sistemas expertos de IA, es decir, desde el diseño del algoritmo o código, hasta la gestión, control y mejora continua de este.

Nada impide garantizar el control del sistema a través de una comisión que se reúna periódicamente, con independencia de su recurrencia –mensual, trimestral, semestral, etc–, mientras que la gestión diaria esté encomendada a responsables técnicos, que reporten, y den cuenta de la vigencia y efectividad de los controles y medidas técnicas-organizativas establecidas desde la misma. Lo importante, una vez más, es lo sustantivo, esto es que las funciones estén o deban estar claras, la incorporación del control y gestión sobre todos los procesos que impliquen la digitalización y el apoyo mediante sistemas técnicos a jueces y fiscales, incluidos los más modernos o revolucionarios, cuya principal arista es precisamente los sistemas expertos de IA. Además de las funciones, también deber ser clara la determinación de quién integra la Comisión y cómo se adoptan sus acuerdos.

El principal cometido o función del órgano de gestión y control debe ser, precisamente, garantizar ciertos principios de naturaleza jurídico-constitucional que deberían inspirar todo el proceso. Es ese órgano de control, integrado por juristas de diversa procedencia que representan lo mejor de distintas profesiones, junto a otros perfiles multidisciplinarios o interdisciplinarios, de naturaleza técnica, quien debe garantizar en la práctica la publicidad y transparencia<sup>6</sup> de todo el proceso –no solo del sistema de IA, empezando por el detalle de quién participa en el órgano de control y quién tiene el encargo o deber de vigilancia y supervisión concreto–, la trazabilidad<sup>7</sup> del código de IA –detalle de todas las acciones que llevan a cabo todos los agentes con poderes de gestión sobre el código o algoritmo, a modo de *backlog* o *audit trail*–, la explicabilidad<sup>8</sup> –hacer fácil lo difícil, hacer comprensible para una persona legla las principales pautas y reglas lógicas que hacen que el algoritmo proyecte esas estimaciones– y auditabilidad –interna y externa, sometido al escrutinio y revisión periódica de ojos expertos ajenos, que puedan certificar eventuales vulnerabilidades o deficiencias en la configuración técnica del sistema–.

Ese es el *prius* necesario para luego acompañar el sistema de determinadas garantías: la impugnabilidad del código, vehiculada a través de los recursos en el caso concreto, como parte del derecho de defensa, del derecho a recurrir<sup>9</sup>, del derecho a tener una segunda oportunidad<sup>10</sup>. En el ámbito de la IA para la valoración de la prueba, las partes deben poder recurrir y protestar las decisiones adoptadas sobre la base de o en apoyo de los sis-

<sup>6</sup> El Parlamento Europeo, al respecto, ha tenido la oportunidad de subrayar que todo sistema de inteligencia artificial debe desarrollarse respetando los principios de transparencia y de responsabilidad algorítmica de modo que los seres humanos puedan comprender sus acciones; y relacionado con lo anterior, indica que, a fin de generar confianza en la inteligencia artificial y permitir el progreso de la misma, los usuarios deben ser conscientes de cómo se utilizan sus datos personales, así como otros datos y datos derivados de sus datos, cuando se comunican o interactúan con un sistema de inteligencia artificial o con humanos que se apoyan en un sistema de inteligencia artificial (Resolución del Parlamento Europeo 2018/2088(INI), 2019, párr. 158).

<sup>7</sup> El Parlamento Europeo, al respecto, ha tenido la oportunidad de subrayar que todo sistema de inteligencia artificial debe desarrollarse respetando los principios de transparencia y de responsabilidad algorítmica de modo que los seres humanos puedan comprender sus acciones; y relacionado con lo anterior, indica que, a fin de generar confianza en la inteligencia artificial y permitir el progreso de la misma, los usuarios deben ser conscientes de cómo se utilizan sus datos personales, así como otros datos y datos derivados de sus datos, cuando se comunican o interactúan con un sistema de inteligencia artificial o con humanos que se apoyan en un sistema de inteligencia artificial (Resolución del Parlamento Europeo 2018/2088(INI), 2019, párr. 158).

<sup>8</sup> El legislador europeo ha destacado la importancia de poder explicar los resultados, los procesos y los valores de los sistemas de inteligencia artificial, de modo que resulten comprensibles para los destinatarios que no tengan conocimientos técnicos y que se les proporcione información significativa, condición necesaria para evaluar la equidad y ganarse la confianza (Resolución del Parlamento Europeo 2018/2088(INI), 2019, párr. 161).

<sup>9</sup> También sobre este extremo se ha manifestado el Parlamento Europeo, que señala que las personas tienen derecho a saber, derecho de recurso y derecho a reparación cuando la inteligencia artificial se utilice para decisiones que les afecten y conlleven un riesgo significativo para los derechos o libertades individuales o puedan menoscabarlos (Resolución del Parlamento Europeo 2018/2088(INI), 2019, párr. 153).

<sup>10</sup> Son muchos los retos sociales derivados de los sistemas de IA predictiva e IAJV de los cuales resultan estimaciones o proyecciones que clasifican a los ciudadanos sobre la base de una elaboración de perfiles. En este ámbito, el Parlamento Europeo ha subrayado la necesidad de que los ciudadanos no sean objeto de discriminación en función de su clasificación y que deban tener derecho a una “segunda oportunidad” (Resolución del Parlamento Europeo 2018/2088(INI), 2019, párr. 12).

temas expertos, alejando cualquier tipo de duda o sombra sobre la imparcialidad del sistema de IA, garantizando y expulsando cualquier atisbo de sesgo discriminatorio, que pueda poner en tela de juicio que el juzgador, formal o sustantivamente, se haya sustentado en un elemento externo distorsionador, que haya condicionado su valoración del caso concreto de forma decisiva, basándose, como decíamos, en un prejuicio de base discriminatoria.

La asunción de esas funciones por parte del órgano de control, en la práctica, descarta el eventual sesgo de los promotores, es decir, de las empresas privadas. El Parlamento Europeo se ha referido a la necesidad de contar con una mano de obra diversificada en todas las ramas del sector de las tecnologías de la información, y con mecanismos de salvaguardia para evitar distorsiones basadas en el género y la edad en los sistemas de IA (Resolución del Parlamento Europeo 2018/2088(INI), 2019, párr. 168). El órgano de control debería, así, asumir un rol de supervisión de procesos y, muy significativamente, de aquellos que, hipotéticamente, pudieran perturbar de forma directa o indirecta el desarrollo de la IA, participando activamente en su diseño y gestión, y trabajando con métodos cuantitativos y cualitativos de prevención de riesgos como los que funcionan, por ejemplo, en el ámbito de la *compliance*, que permiten la detección de casos en los que los sistemas de IA pueden haber sido manipulados de forma accidental o malintencionada.

#### IV. REFLEXIONES FINALES

**Primera.-** A lo largo de este trabajo hemos observado algunos usos razonables acerca del empleo de la IA en la Administración de Justicia. Una administración que, como Tercer Poder, se encuentra ante la oportunidad o el precipicio, y que está condenada a ser digital o a su desaparición (Perea González, 2020).

**Segunda.-** El uso de la IA en la justicia entraña tantos o más riesgos que oportunidades, especialmente para los derechos, principios y garantías jurídico-constitucionales que reúne el proceso y, muy especialmente, el proceso penal. La pérdida del control y transparencia sobre el algoritmo puede derivar en una afectación de derechos fundamentales de la ciudadanía, tales como el derecho de defensa, la intimidad, la protección de los datos personales, la imparcialidad o independencia judicial o el principio de publicidad y transparencia de las actuaciones judiciales. Por límites espaciales obvios no hemos podido abordar aquí estos retos, si bien hemos tenido ocasión de formular un análisis en detalle en otro foro (Simón Castellano, 2021b).

**Tercera.-** La valoración de la prueba sin lugar a duda es uno de los ámbitos en los que el empleo de IA tiene más posibilidades y resultaría más útil, al contribuir a objetivar la libre valoración de la prueba del juzgador. Respecto a la prueba documental, las ventajas son indubitadas. Los sistemas de IA pueden comprender el significado de los textos, sea cual sea su idioma, y procesarlo ofreciendo una visión que ayudará a complementar la interpretación y valoración humana del juzgador, que ahora puede apoyarse en estas modernas técnicas. En relación con la prueba pericial, los sistemas de IA pueden contribuir a valorar con acierto y profundidad el currículo del perito y automatizar la aplicación al caso concreto de los criterios Daubert. Finalmente, por lo que se refiere a la valoración de las personas, los sistemas de IA pueden contribuir en aquel aspecto en el que la habilidad humana muestra su faceta más débil, que es, precisamente, las corroboraciones, al poder comprobar o chequear a tiempo real la compatibilidad de la declaración de testigos o partes con el resto de las pruebas admitidas y practicadas en sede judicial.

**Cuarta.-** Como consecuencia y reacción frente a la suma de la segunda –riesgos inherentes desde la óptica de los derechos y principios constitucionales– y tercera –notables oportunidades para la valoración racional y objetiva de la prueba– conclusión, resulta fundamental el papel que debe desempeñar el organismo de control como principal garante de la aplicación práctica de ciertos principios que, de lo contrario, quedarían en papel mojado. Nos referimos a los principios de publicidad, transparencia, accesibilidad, auditabilidad, explicabilidad y trazabilidad del código o algoritmo, a los que habría que sumar, como novedad y solo por lo que se refiere al empleo de la IA en sede judicial, el principio de jurisdiccionalidad y legalidad desde el diseño y por defecto. Un organismo que debería incluir la visión de los principales actores del proceso judicial, y cuya participación y control debería viajar desde la inceptión, es decir, desde el diseño y la primera idea, hasta su implementación y gestión diaria, pasando por la necesaria revisión y mejora continua del sistema. 📌

#### REFERENCIAS

Asociación Profesional de la Magistratura [@juecesAPM] (11 de diciembre de 2020). Comisión Ministerial de Administración Digital del @justiciagob sin ningún juez, ni fiscal, ni LAJ, ni abogado, ni funcionario. Gente que no conoce un procedimiento, no lo ha visto, ni sabe los problemas diarios de los juzgados. Fracaso asegurado @Jccampm [Enlace e imagen adjuntos] [Tweet]. Twitter. <https://twitter.com/juecesAPM/status/1337312803821133824>

- Babuta, A., Oswald, M., & Rinik, C. (2018). *Machine Learning Algorithms and Police Decision-Making: Legal, Ethical and Regulatory Challenges* (Whitehall Reports, Vol 3, No. 18). Royal United Services Institute for Defense and Security Studies. <https://rusi.org/publication/whitehall-reports/machine-learning-algorithms-and-police-decision-making-legal-ethical>
- Balbuena Pérez, D. E. (2017). *Derecho penal, neurociencia y libertad: una aproximación teórica al replanteamiento de la responsabilidad penal a propósito de las aportaciones neurocientíficas sobre el proceso mental para formación de la voluntad* (3ra ed.). Marben.
- Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671-732. <http://dx.doi.org/10.15779/Z38BG31>
- Belfiore, R. (2013). The Protection of Personal Data Processed Within the Framework of Police and Judicial Cooperation in Criminal Matters. En S. Ruggeri (ed.), *Transnational Inquiries and the Protection of Fundamental Rights in Criminal Proceedings* (pp. 355-370). Springer.
- Berrocal Lanzarot, A. I. (2017). *Derecho de supresión de datos o derecho al olvido*. Reus.
- Boix Palop, A. (2020). Los algoritmos son reglamentos: La necesidad de extender las garantías propias de las normas reglamentarias a los programas empleados por la administración para la adopción de decisiones. *Revista de Derecho Público. Teoría y Método*, 1, 223-269. [https://doi.org/10.37417/RPD/vol\\_1\\_2020\\_33](https://doi.org/10.37417/RPD/vol_1_2020_33)
- Bromby, M. C., & Hall, M. J. (2002). The Development and Rapid Evolution of the Knowledge Model of ADVOKATE: An Advisory System to Assess the Credibility of Eyewitness Testimony. En T. J. M. Bench-Capon, A. Daskalopulu, & R. G. F. Winkels (eds.), *Legal Knowledge and Information Systems. The Fifteenth Annual Conference* (pp. 143-152). IOS Press.
- Cabellos Espiérrez, M. A. (2018). Opinar, enaltecer, humillar: respuesta penal e interpretación constitucionalmente adecuada en el tiempo de las redes sociales. *Revista Española de Derecho Constitucional*, 112, 45-86.
- Calders, T., Kamiran, F., & Pechenikiy, M. (2009). Building classifiers with independency constraints. *2009 IEEE International Conference on Data Mining Workshops*, 13-18. <https://doi.org/10.1109/ICDMW.2009.83>
- Cerrillo i Martínez, A. (2019). El impacto de la inteligencia artificial en el derecho administrativo ¿nuevos conceptos para nuevas realidades técnicas?. *Revista General de Derecho Administrativo*, (50).
- Cerrillo i Martínez, A., & Peguera Poch, M. (2020). *Retos jurídicos de la inteligencia artificial*. Aranzadi.
- Chayes, A. (1976). The Role of the Judge in Public Law Litigation. *Harvard Law Review*, 89(7), 1281-1316. <https://doi.org/10.2307/1340256>
- Citron, D. K., & Pasquale, F. (2014). The scored society: due process for automated predictions. *Washington Law Review*, 89, 1-33.
- Cobacho López, A. (2019). Reflexiones en torno a la última actualización del derecho al olvido digital. *Revista de Derecho Político*, (1)104, 198-227. <https://doi.org/10.5944/rdp.104.2019.24313>
- Cotino Hueso, L. (2019). Riesgos e impactos del Big Data, la inteligencia artificial y la robótica. enfoques, modelos y principios de la respuesta del derecho. *Revista General de Derecho Administrativo*, 50.
- (2020). SyRI, ¿a quién sanciono? Garantías frente al uso de inteligencia artificial y decisiones automatizadas en el sector público y la sentencia holandesa de febrero de 2020. *La Ley Privacidad*, (4).
- Council of Bars and Law Societies of Europe (2020). *Considerations on the Legal Aspects of Artificial Intelligence*. CCBE.
- Crawford, K. & Schultz, J. (2014). Big data and due process: Toward a framework to redress predictive privacy harms. *Boston College Law Review*, 55(1), 93-128.
- Creative Destruction Lab (22 de septiembre de 2017). Remarks from Daniel Kahneman [video]. Youtube. [https://www.youtube.com/watch?v=gbj\\_NsgNe7A](https://www.youtube.com/watch?v=gbj_NsgNe7A)
- De Asís Roig, R. F. (2018). Robótica, inteligencia artificial y derecho. *Revista de privacidad y derecho digital*, 3(10), 27-77.
- De Miguel Beriain, I. (2018). Does the use of risk assessments in sentences respect the right to due process? A critical analysis of the Wisconsin v. Loomis ruling. *Law, Probability and Risk*, 1(17), 45-53. <https://doi.org/10.1093/lpr/mgy001>

- De Souza, K. (2018). *Delivering Artificial Intelligence in Government: Challenges and Opportunities*. IBM Center for Business and Government. <https://www.businessofgovernment.org/report/delivering-artificial-intelligence-government-challenges-and-opportunities>
- Dumortier, F. (2010). Facebook and Risks of 'De-contextualization' of Information. En S. Gutwirth, Y. Poulet, & P. De Hert (eds.), *Data Protection in a Profiled World* (pp. 119-138). Springer.
- Edwards, L., & Veale, M. (2017). Slave to the Algorithm: Why a Right to an Explanation Is Probably Not the Remedy You Are Looking for. *Duke Law & Technology Review*, 16, 18-84.
- Eggers, W. D., Schatsky, D., & Viechnicki, P. (2018). *AI-augmented government. Using cognitive technologies to redesign public sector work*. DeLoitte University Press.
- Ekman, P. (2011). *Cómo detectar mentiras: una guía para utilizar en el trabajo, la política y la pareja*. Paidós.
- Ercilla García, J. (15 de diciembre de 2020). Automatización de la justicia: un ejemplo práctico. *Diario La Ley*. [https://diariolaley.laleynext.es/Content/Documento.aspx?params=H4slAAAAAAEAMtMSbF1CTEAAmNDCzMjI7Wy1KLizPw8WyMDIwNDQ0MLtbz8INQQF2fb0ryU1LTMvNQkJLMTeqX\\_OSQyoJU27TEenOJUtdSk\\_Pxs-FJPIYSYAALhfl5tjAAAAWKE](https://diariolaley.laleynext.es/Content/Documento.aspx?params=H4slAAAAAAEAMtMSbF1CTEAAmNDCzMjI7Wy1KLizPw8WyMDIwNDQ0MLtbz8INQQF2fb0ryU1LTMvNQkJLMTeqX_OSQyoJU27TEenOJUtdSk_Pxs-FJPIYSYAALhfl5tjAAAAWKE)
- Executive Office of the President National Science & Technology Council Committee on Technology. (2016). *Preparing for the Future of Artificial Intelligence*. [https://obamawhitehouse.archives.gov/sites/default/files/whitehouse\\_files/microsites/ostp/NSTC/preparing\\_for\\_the\\_future\\_of\\_ai.pdf](https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf)
- Fernández Hernández, C. B. (2020a, 7 de mayo). El Consejo de la Abogacía Europea analiza los efectos de la aplicación de la Inteligencia Artificial en el ámbito jurídico. *Diario La Ley*. <https://diariolaley.laleynext.es/dll/2020/05/07/el-consejo-de-la-abogacia-europea-analiza-los-efectos-de-la-aplicacion-de-la-inteligencia-artificial-en-el-ambito-juridico>
- (2020b, 24 de julio). Informe para el Parlamento Europeo sobre el uso de la inteligencia artificial en los ámbitos policial y judicial. *Diario La Ley*. <https://diariolaley.laleynext.es/dll/2020/07/24/informe-para-el-parlamento-europeo-sobre-el-uso-de-la-inteligencia-artificial-en-los-ambitos-policial-y-judicial>
- (2020c, 13 de diciembre). Primera sentencia europea que declara ilegal un algoritmo de evaluación de características personales de los ciudadanos. *Diario La Ley*. <https://diariolaley.laleynext.es/Content/Documento.aspx?params=H4slAAAAAAEAMtMSbH1czUwMDAY-NDa3NDJUK0stKs7Mz7M1MjACC6rl5aekhrG425bmpaSmZealpoCUZKZVuuQnh1QWpNqmJeYUp6qlJuXnZ6OYFA8zAQCFsDkrYwAAAA=WKE>
- Ferrer Beltrán, J. (2002). *Prueba y verdad en el derecho*. Marcial Pons.
- (2007). *La valoración racional de la prueba*. Marcial Pons.
- Frey, C. B., & Osborne, M. A. (2017). The future of employment: how susceptible are jobs to computerization?. *Technological forecasting and social change*, 14, 254-280. <https://doi.org/10.1016/j.techfore.2016.08.019>
- Gómez Manresa, F., & Fernández Salmerón, M. (2019). *Modernización digital en la administración de la justicia*. Aranzadi.
- González Fuster, G. (2020). *Artificial Intelligence and Law Enforcement - Impact on Fundamental Rights*. European Parliament Think Tank Study. [https://www.europarl.europa.eu/thinktank/en/document.html?reference=IPOL\\_STU\(2020\)656295](https://www.europarl.europa.eu/thinktank/en/document.html?reference=IPOL_STU(2020)656295)
- Harari, Y. N. (2018). *21 lecciones para el siglo XXI*. Debate.
- HarvardCPL (14 de febrero de 2014). Daniel Kahneman, in conversation with Cass Sunstein [video]. Youtube. <https://www.youtube.com/watch?v=NUvoEXGD7bE>.
- IT-POL (19 de junio de 2019). Danish DPA approves Automated Facial Recognition. *European Digital Rights Initiative*. <https://edri.org/danish-dpa-approves-automated-facial-recognition/>
- Julià Pijoan, M. (2020). *Proceso penal y (neuro)ciencia: una interacción desorientada. Una reflexión acerca de la neuropredicción*. Marcial Pons.
- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press.
- Kaplan, J. (2017). *Inteligencia artificial. Lo que todo el mundo debe saber*. Teell Editorial.
- Libre valoración de la prueba (2020). En *Diccionario panhispánico del español jurídico*. <https://>

- dpej.rae.es/lema/libre-valoraci%C3%B3n-de-la-prueba
- Martínez Garay, L. (2018). Peligrosidad, algoritmos y due process: el caso State v Loomis. *Revista de Derecho Penal y Criminología*, (20), 485-502. <https://doi.org/10.5944/rdpc.20.2018.26484>
- Martínez Martínez, R. (2019). Inteligencia artificial desde el diseño. Retos y estrategias para el cumplimiento normativo. *Revista Catalana de Dret Públic*, (58), 64-81. <http://dx.doi.org/10.2436/rcdp.i58.2019.3317>
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data. La revolución de los datos masivos*. Turner.
- Miró Llinares, F. (2018). Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots. *Revista de Derecho Penal y Criminología*, 20, 87-130.
- Myers, D. G. (2002). *Intuición. El poder y el peligro del sexto sentido*. Paidós.
- Nieva Fenoll, J. (2010). *La valoración de la prueba*. Marcial Pons.
- (2018). *Inteligencia artificial y proceso judicial*. Marcial Pons.
- Perea González, A. (24 de junio de 2020). Tercer Poder: la Justicia ante la oportunidad o el precipicio. *Diario el Mundo*. <https://www.elmundo.es/opinion/columnistas/2020/06/24/5ef327c921efa062678b45f8.html>
- Ponce Solé, J. (2019). Inteligencia artificial, Derecho administrativo y reserva de humanidad: algoritmos y procedimiento administrativo debido tecnológico. *Revista General de Derecho Administrativo*, 50.
- Posner, E. A. (2011). *Cómo deciden los jueces* (V. Roca Pérez, trad.). Marcial Pons.
- Richardson, R., Schultz, J. M., & Crawford, K. (2019). Dirty Data, Bad Predictions: how Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice. *New York University Law Review*, (94), 15-55.
- Romeo Casabona, C. M. (2018). Riesgo, procedimientos actuariales basados en inteligencia artificial y medidas de seguridad. *Revista de Derecho, Empresa y Sociedad (R.E.D.S.)*, (13), 39-55.
- Sierra, S. (2020). Inteligencia artificial y justicia administrativa: una aproximación desde la teoría del control de la Administración Pública. *Revista General de Derecho Administrativo*, 53.
- Simon Castellano, P. (2012). *El régimen constitucional del derecho al olvido digital*. Tirant lo Blanch.
- (2015). *El reconocimiento del derecho al olvido digital en España y en la UE*. Wolters Kluwer - Bosch.
- (2021a). Inteligencia Artificial y Administración de Justicia: ¿Quo vadis, justitia?. *Revista Internet, Derecho y Política (IDP)*, 33.
- (2021b). *Justicia cautelar e inteligencia artificial. La alternativa a los atávicos heurísticos judiciales*. J. M. Bosch Editor.
- Souza de Menezes, C., & Agustina, J. (2020). Big Data, Inteligencia Artificial y policía predictiva. Bases para una adecuada regulación legal que respete los derechos fundamentales. En D. Dupuy & J.G. Corvalán (dirs.), & M. Kiefer (coord.), *Ciberdelitos III inteligencia artificial, automatización, algoritmos y predicciones en el derecho penal y procesal penal* (pp. 137-182). Editorial B de F.
- Taruffo, M. (2002). *La prueba de los hechos* (J. Ferrer Beltrán, trad.). Trotta. (Obra original publicada en 1992).
- Valero Torrijos, J. (2019). Las garantías jurídicas de la inteligencia artificial en la actividad administrativa desde la perspectiva de la buena administración. *Revista Catalana de Dret Públic*, (58), 82-96. <http://dx.doi.org/10.2436/rcdp.i58.2019.3307>
- Vázquez Rojas, C. (2015). *De la prueba científica a la prueba pericial*. Marcial Pons.
- Zarsky, T. (2016). The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making. *Science, Technology, & Human Values*, 41(1), 118-132.

#### LEGISLACIÓN, JURISPRUDENCIA Y OTROS DOCUMENTOS LEGALES

Constitución Española [C.E.], B.O.E. n. 311, Dec. 29, 1978 (Esp.).

Daubert c. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993).

Estado de Wisconsin c. Loomis, 881 N.W.2d 749 (Wis. 2016).

European Commission for the Efficiency of Justice [CEPEJ], European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment, adopted on the 31st plenary (Dec. 3-4, 2018) (UE).

Fed. R. Evid. 702 (amended Dec. 1, 2020) (EE. UU.).

General Electric Co. c. Joiner, 522 U.S. 136 (1997).

Informe AEPD 385661/2017 de marzo relativo al anteproyecto de ley de sistema de alertas tempranas de la Comunidad Valenciana, Agencia Española de Protección de Datos, de 13 de marzo de 2018 (Esp.).

Kumho Tire Co. c. Carmichael, 526 U.S. 137 (1999).

Ley de Enjuiciamiento Civil (B.O.E. 2000, 1) (Esp.).

Ley de Inspección General de Servicios y del sistema de alertas para la prevención de malas prácticas en la Administración de la Generalitat y su sector público instrumental (B.O.E. 2018, 22) (Esp.).

Orden JUS/1179/2020, de 4 de diciembre, por la que se modifica la Orden JUS/1362/2016, de 3 de agosto, por la que se crea la Comisión Ministerial de Administración Digital del Ministerio de Justicia y se regula su composición y funciones (B.O.E. 2020, núm. 322) (Esp.).

Rb Dan Haag 5 februari 2020, NJ 2020, 368 m.nt. van E.J. Dommering (NJCM et al./Países Bajos) (Países Bajos).

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

Resolución del Parlamento Europeo, de 12 de febrero de 2019, sobre una política industrial global europea en materia de inteligencia artificial y robótica (2018/2088(INI)), 2019 O.J. (C 449) 39.

Resolución del Parlamento Europeo, 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 (2020/2012(INL)), 2020 O.J. (C 404) 63.