

BOLETÍN

INTELIGENCIA ARTIFICIAL

VOLUMEN VII

Agosto
2026

Juezas y Jueces
para la Democracia



ARTICULOS

INTELIGENCIA ARTIFICIAL Y ELABORACIÓN DE RESOLUCIONES JUDICIALES

Andrés Montalbán Losada

LA INDEPENDENCIA JUDICIAL EN LA ERA DE LA INTELIGENCIA ARTIFICIAL

Sandra María Piñeiro Vilas

INTELIGENCIA ARTIFICIAL GENERATIVA

Miguel Ángel Serrano

BOLETÍN

INTELIGENCIA ARTIFICIAL

VOLUMEN VII

Dirección

Diego Gutiérrez Alonso
María Gavilán Rubio

Coordinación

Fátima Mateos Hernández

Diseño y maquetación

Emi Ramírez

Imágenes

Freepick, Unsplash

Edita

Juezas y Jueces para
la Democracia en Madrid

Juezas y Jueces
para la Democracia

ARTÍCULOS

**Inteligencia Artificial y elaboración
de resoluciones judiciales.
Experiencia práctica, herramientas
disponibles y límites institucionales** 5
Andrés Montalbán Losada

**La independencia judicial en
la era de la Inteligencia Artificial.
Sesgos, opacidad y delegación
cognitiva en el oficio de juzgar** 14
Sandra María Piñeiro Vilas

**Inteligencia Artificial generativa.
Panorama desde un escritorio** 48
Miguel Ángel Serrano

Editorial

Inteligencia artificial y Justicia: innovar sin renunciar al juicio humano

La inteligencia artificial ha dejado de ser una promesa de futuro para convertirse en una realidad presente en todos los ámbitos de la actividad profesional. La Administración de Justicia no constituye una excepción. Herramientas capaces de analizar documentación masiva, localizar jurisprudencia, sintetizar información compleja o asistir en la redacción de textos jurídicos han irrumpido en un ecosistema tradicionalmente caracterizado por el rigor, la prudencia y la responsabilidad institucional. La cuestión ya no es si la inteligencia artificial formará parte del trabajo de jueces, magistrados, letrados y operadores jurídicos, sino cómo integrarla sin comprometer los principios que sostienen el Estado de Derecho.

Los tres trabajos que integran este número abordan esta cuestión desde perspectivas complementarias que permiten al lector obtener una visión amplia, crítica y equilibrada de uno de los mayores desafíos de nuestro tiempo. Lejos de ofrecer una mirada simplista, los autores coinciden en una idea fundamental: la tecnología puede transformar profundamente la Justicia, pero no puede sustituir aquello que constituye su esencia, esto es, la capacidad humana de juzgar con independencia, responsabilidad y sometimiento exclusivo al Derecho.

Andrés Montalbán Losada aporta una perspectiva eminentemente práctica sobre la utilización de la inteligencia artificial en la elaboración de resoluciones judiciales. Desde la experiencia cotidiana de quien ejerce la función jurisdiccional, examina el marco establecido por la Instrucción 2/2026 del Consejo General del Poder Judicial y distingue con claridad entre los usos legítimos de la inteligencia artificial como instrumento de apoyo y aquellos que resultarían incompatibles con la función de juzgar. Su análisis pone de relieve que el verdadero reto no consiste en impedir la innovación tecnológica, sino en preservar la autoría intelectual de la decisión judicial, evitando que la asistencia tecnológica se transforme, de forma inadvertida, en una delegación del juicio jurisdiccional.

Por su parte, Sandra María Piñeiro Vilas sitúa el debate en una dimensión institucional y constitucional de gran profundidad. Su reflexión sobre la independencia judicial en la era de la inteligencia artificial advierte acerca de riesgos que trascienden la mera eficiencia de las herramientas digitales: los sesgos algorítmicos, la opacidad tecnológica y la llamada delegación cognitiva. La autora propone ampliar la concepción clásica de la independencia judicial incorporando la idea de independencia cognitiva, entendida como la capacidad efectiva del juez para formar un criterio propio, comprender las razones de una recomendación tecnológica y apartarse de ella cuando el caso concreto así lo exija. En un contexto donde los algoritmos pueden influir silenciosamente en la selección de información y en la construcción de las decisiones, esta reflexión resulta especialmente pertinente.

El tercer artículo, firmado por Miguel Ángel Serrano, amplía el foco más allá del ámbito estrictamente judicial y examina las implicaciones sociales, económicas, culturales y políticas de la inteligencia artificial generativa. Su mirada crítica invita a reflexionar sobre

cuestiones que afectan directamente al futuro de nuestras sociedades: la protección de la propiedad intelectual, el impacto sobre las profesiones creativas, la concentración de poder tecnológico, la sostenibilidad económica de los modelos actuales y los riesgos derivados de una dependencia excesiva de sistemas automatizados. Desde esta perspectiva, la inteligencia artificial aparece no solo como una innovación tecnológica, sino como un fenómeno capaz de redefinir relaciones de poder, modelos productivos y espacios de libertad individual.

La lectura conjunta de estos tres trabajos permite advertir una conclusión compartida. La inteligencia artificial ofrece oportunidades indudables para mejorar la calidad y la eficiencia de la Administración de Justicia. Puede liberar recursos, agilizar tareas repetitivas y facilitar el acceso al conocimiento jurídico. Sin embargo, sus beneficios solo serán plenamente compatibles con un sistema democrático si se desarrollan dentro de un marco sólido de garantías, transparencia, supervisión y formación.

La Justicia no es únicamente una actividad técnica. Es también una función constitucional que exige responsabilidad, motivación y control. Por ello, la incorporación de nuevas herramientas debe orientarse a reforzar las capacidades del profesional jurídico, nunca a sustituirlas. La tecnología puede asistir al juez; no puede asumir su responsabilidad. Puede ayudar a encontrar información; no puede decidir qué es justo. Puede generar textos; no puede reemplazar el razonamiento que les otorga legitimidad.

Este número invita, en definitiva, a participar en un debate imprescindible. Un debate que no enfrenta a partidarios y detractores de la innovación, sino que exige encontrar el difícil equilibrio entre progreso tecnológico y garantías jurídicas. Porque, en última instancia, la verdadera cuestión no es cuánto puede hacer la inteligencia artificial, sino qué estamos dispuestos a preservar como espacio irreductiblemente humano dentro de la función de juzgar.

Conecta con JIpD

¡SIGUENOS!



@jpdemocracia.bsky.social



@JpDemocracia

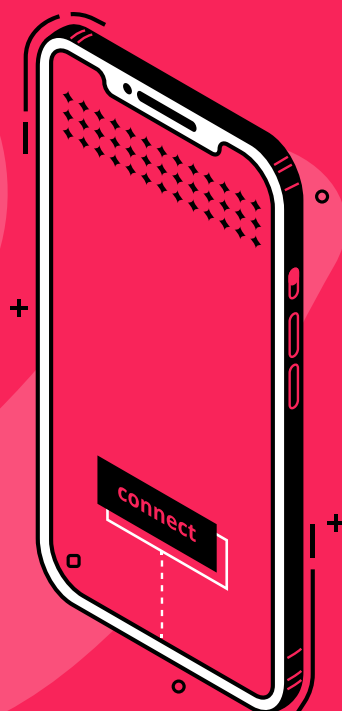


@jjpdemocracia



Juezas y Jueces
para la Democracia

JIpD



Inteligencia Artificial y elaboración de resoluciones judiciales

Experiencia práctica, herramientas
disponibles y límites institucionales

Andrés Montalbán Losada

Magistrado. Sección de lo contencioso administrativo del Tribunal de Instancia de Murcia

I. PLANTEAMIENTO

La pregunta pertinente ya no es si la inteligencia artificial entrará en los juzgados, porque ha entrado, sino bajo qué condiciones puede el juez servirse de ella para elaborar sus resoluciones sin ceder la autoría intelectual de la decisión. El artículo 117 de la Constitución no contiene una cláusula puramente orgánica: al reservar en exclusiva a juzgados y tribunales la potestad de juzgar, configura una forma singular de ejercicio del poder público, personal, motivada, independiente, responsable y jurídicamente controlable. La amenaza sería no es la imagen tópica de una máquina dictando sentencias; es bastante más discreta: que el juez conserve la firma y vaya perdiendo, casi sin advertirlo, la autoría intelectual de la resolución, que seguiría siendo formalmente judicial mientras viene materialmente prefigurada por una arquitectura técnica que selecciona fuentes, jerarquiza argumentos y normaliza respuestas.

Estas líneas no son una resolución ni un tratado. Nacen de la práctica diaria y de una formación en buena medida autodidacta, y se dirigen a otros compañeros con una pretensión deliberadamente modesta y práctica: ofrecer una lectura clara de los límites que fija la Instrucción 2/2026 del Consejo General del Poder Judicial, precisar dónde se sitúa hoy la verdadera frontera en relación con las herramientas realmente disponibles, y describir un método de trabajo que permite aprovechar la asistencia tecnológica sin abdicar de la función jurisdiccional. Quien escribe no lo hace desde la teoría, sino desde el uso cotidiano de estas herramientas; y si algo ha aprendido es que conviene hablar de ellas con naturalidad, sin reverencia tecnológica ni recelo corporativo.

II. EL MARCO: LA INSTRUCCIÓN 2/2026 Y LA REGLA DE CIERRE

La Instrucción 2/2026 (Acuerdo de 28 de enero de 2026, del Pleno del Consejo General del Poder Judicial) admite el empleo de sistemas de inteligencia artificial como instrumentos de apoyo o asistencia, pero excluye que puedan operar autónomamente para adoptar decisiones judiciales, valorar hechos o pruebas, o interpretar y aplicar el Derecho. Sus principios —control humano efectivo, no sustitución, responsabilidad judicial, independencia, prevención de sesgos, proporcionalidad y uso limitado— son una concreción institucional de las garantías del artículo 117 CE.

De ella deriva, para el juez de a pie, una regla de cierre que conviene retener con precisión: solo cabe utilizar sistemas facilitados por las Administraciones competentes en materia de justicia o por el propio Consejo, sometidos a control de calidad y auditoría, y únicamente como instrumentos de apoyo. Entre los usos permitidos figuran la búsqueda de información jurídica, el análisis, la clasificación y la estructuración de documentos, la elaboración de esquemas, resúmenes o borradores de trabajo interno y el apoyo a tareas organizativas. Pero los borradores de resolución generados mediante IA solo pueden emplearse en esos términos: exigen revisión y validación personal, completa y crítica por el juez; no tienen la consideración de decisiones automatizadas; y han de poder ser libremente modificados antes de su validación como resolución. Queda expresamente prohibida la sustitución, automatización o delegación de la toma de decisiones, de la valoración de hechos o pruebas y de la interpretación y aplicación del Derecho.

El dato decisivo para el práctico es, por tanto, que la utilidad tecnológica no equivale a legitimidad. Que una herramienta sea potente, o que llegue cómodamente al escritorio del

juez, no la convierte en jurídicamente admisible: lo que importa es su procedencia institucional, su auditabilidad y la garantía de que los datos del expediente no salgan en momento alguno del perímetro institucional.

III. DOS «COPILOT» DISTINTOS: CUÁL PUEDE UTILIZARSE Y POR QUÉ

Es aquí donde se sitúa el problema más concreto y peor resuelto, y conviene empezar deshaciendo un equívoco muy extendido: cuando se habla del «Copilot» de los jueces no se habla de una sola herramienta, sino de dos, de naturaleza y régimen bien distintos. Confundirlas es la fuente de casi todos los malentendidos.

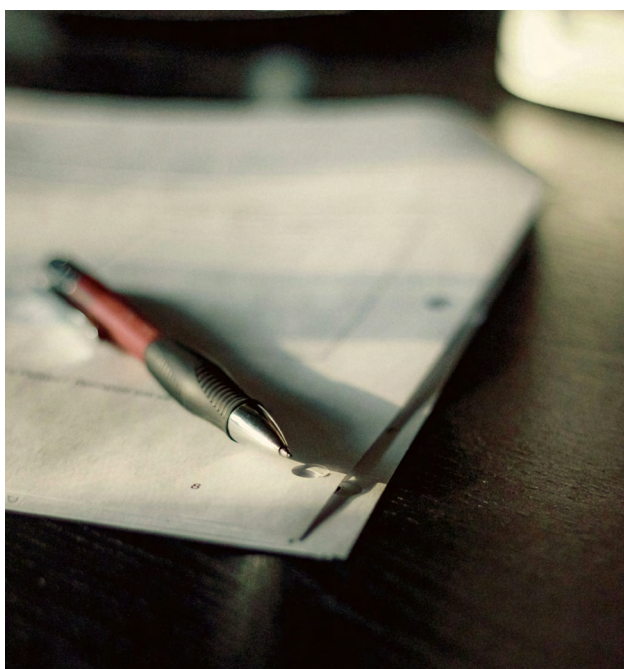
El primero es el Copilot del Consejo General del Poder Judicial

Es el accesible a través de la página web del Consejo y de nuestra intranet, y llega por un contrato del propio Consejo que incorpora un paquete en el que conviven Outlook, Copilot y OneDrive. El problema,

según ha reconocido el Consejo al contestar a las consultas que le han sido planteadas, es que ese contrato no detalló las condiciones en que debía suministrarse el Copilot; y, al no haberlo hecho, la herramienta no está auditada. De ahí extrae dos consecuencias: que no es susceptible de utilizarse subiendo documentos que incorporen datos protegidos, y que tampoco constituye, en sentido propio, una inteligencia artificial «facilitada» en los términos que exige la regla de cierre de la Instrucción. El mero hecho de que una herramienta alcance el equipo del juez por la vía de un licenciamiento corporativo no la transforma en un sistema institucionalmente validado.

El segundo es el Copilot del Ministerio de Justicia

Es el que el Ministerio facilita en aquellos territorios cuya Administración de Justicia depende de él, y no de las comunidades autónomas con competencias transferidas. Este viene contratado dentro de Horus — la herramienta de tramitación y gestión del expediente judicial electrónico que el Ministerio pone a disposición de los tribunales de instancia, las audiencias provinciales y



“El mero hecho de que una herramienta alcance el equipo del juez por la vía de un licenciamiento corporativo no la transforma en un sistema institucionalmente validado”

los tribunales superiores de justicia de su territorio, vinculada al sistema de gestión procesal Atenea, sucesor del anterior Minerva, con extensión a la Audiencia Nacional y al Tribunal Supremo— e integrado en el propio entorno que da acceso a los procedimientos. Según pude constatar en una actividad celebrada a principios de junio de 2026 en la sede de la formación continua del Consejo General del Poder Judicial, y conforme expuso don Aitor Cubo, Director General de Transformación Digital de la Administración de Justicia, este Copilot sí está auditado; en los términos en que está contratado, no puede formarse ni entrenarse con los datos que se le trasladan, que tampoco salen del sistema; y tiene su sede en territorio de la Unión Europea, y no fuera de ella, como exige la regulación comunitaria. Cumple, por tanto —en principio y según el propio Ministerio—, las condiciones que la Instrucción 2/2026 reclama; y, siendo el Ministerio administración competente para facilitar sistemas de inteligencia artificial, se trata de una herramienta institucionalmente facilitada en sentido estricto.

La consecuencia práctica debe quedar nítida: de los dos, solo el segundo —el Copilot del Ministerio de Justicia, integrado en Horus— puede emplearse para trabajar con los datos del expediente. Así se desprende tanto de lo que el Consejo General del Poder Judicial ha respondido a las consultas que se le han planteado como de las propias manifestaciones del Ministerio de Justicia, vertidas a través de su personal autorizado y, señaladamente, de su Director General de Transformación Digital de la Administración de Justicia. La frontera que traza la Instrucción no discurre, pues, entre marcas ni proveedores —ambas herramientas son «Copilot», ambas de Microsoft—, sino entre el sistema institucionalmente facilitado, auditado, no entrenado con el dato y con sede en la Unión Europea, y el que no ofrece esas garantías. Dejo constancia de ello como testimonio de lo expuesto en aquella sesión; conviene, en todo caso, conocer las con-

diciones concretas de uso a través de la documentación oficial del Ministerio.

A partir de esta misma distinción se abre un segundo debate, todavía no cerrado, sobre el alcance de la anonimización dentro del entorno facilitado. Es discutible, pero entiendo que sería perfectamente defendible que el Copilot del Ministerio de Justicia pueda utilizarse sin necesidad de anonimizar previamente el material. El argumento es sencillo: si el propio Ministerio sostiene que la herramienta no puede emplear los datos suministrados para enriquecerse y que esos datos no salen del sistema, nos hallaríamos ante un contrato de naturaleza análoga al que ya rige para el Outlook corporativo —también de Microsoft— o para el OneDrive, que es precisamente el lugar donde hoy se guardan las resoluciones de jueces y magistrados para evitar el trasiego de pinchos y memorias USB, fuente nada desdeñable de pérdidas y de que datos protegidos acaben en manos de terceros. Si admitimos, sin reparo, alojar el expediente y las resoluciones en OneDrive o tramitar información sensible por correo electrónico al amparo de un contrato de confidencialidad, no se ve por qué habría de exigirse una anonimización previa para un Copilot sujeto a esa misma lógica contractual de confidencialidad y no entrenamiento, con sede en la Unión Europea.

Conviene, no obstante, manejar esta tesis con cautela. Una cosa es que el contrato de confidencialidad y no enriquecimiento module el riesgo de fuga o de reutilización del dato, y otra distinta que desplace sin más el conjunto de garantías que la normativa de protección de datos y la propia Instrucción anudan al tratamiento de información sensible en sede judicial. Mientras la cuestión no quede expresamente resuelta por el Consejo o por la autoridad de supervisión, la prudencia aconseja, cuando menos, confirmar las condiciones concretas del entorno facilitado —perímetro cerrado, no entrenamiento y base jurídica del tratamiento— y,

en la duda, mantener la anonimización como regla de seguridad. Lo que en ningún caso parece sostenible es trasladar esa eventual relajación a los sistemas no facilitados: respecto de ellos, la subida de datos protegidos sigue vedada.

IV. UNA METODOLOGÍA RESPONSABLE: ASISTIR SIN SUSTITUIR

Dentro de ese marco, el uso legítimo de la IA en la elaboración de resoluciones puede ordenarse en torno a un principio sencillo: el reparto de autonomía. Hay tareas que son síntesis y sistematización fiel del material aportado y pueden ejecutarse con autonomía, porque no deciden; y hay otra que permanece subordinada a la dirección del juzgador, porque en ella reside el juicio.

Tareas de síntesis fiel (con autonomía)

Pertenecen a este grupo la redacción de los fundamentos que resumen la posición de cada parte —objeto litigioso, suplico y fundamentos de la pretensión actora; motivos de oposición de la demandada—; el inventario y la sistematización de la prueba documental y pericial, con su contenido esencial, su clasificación temática, los patrones de corroboración, las contradicciones y las lagunas, sin atribuir fuerza probatoria ni credibilidad, que corresponden al juzgador conforme a la sana crítica; y el resumen analítico de sentencias, con separación de la ratio decidendi y los obiter dicta y la indicación de su valor jurisprudencial. Para todas ellas he sistematizado modelos de instrucción específicos para la jurisdicción civil y para la contencioso-administrativa, que delimitan el rol del asistente, el material que debe aportarse, la estructura de salida y unas reglas de rigor.

Tarea de dar forma, subordinada a la dirección del juez (el uso decisivo)

Una vez que el juzgador ha analizado la prueba practicada, los escritos de demanda y contestación y el Derecho aplicable, y ha formado su convicción, dicta de manera rápida y sucinta el sentido del fallo y las razones por las que la pretensión prospera o no; y el asistente se limita a darles forma: escribe, en lenguaje solemne, narrativo y diáfano, claro y dirigido al ciudadano pero propio de una resolución



de alto tribunal, aquello que el juez ha decidido y dictado. La decisión, la dirección jurídica y el razonamiento siguen siendo del juzgador; el asistente no añade argumentos propios, salvo que se le pida desarrollar o reforzar alguno. No decide: redacta lo ya decidido.

En este punto, la verdadera utilidad es la velocidad. El Copilot puede resumir los escritos de parte; y, sobre ese resumen, el juez introduce su propia valoración de la prueba, su análisis del Derecho que entiende aplicable y la subsunción del supuesto concreto en la norma. Lo decisivo es que el juzgador puede dictar todas esas ideas de corrido, sin necesidad de darles una forma escrita acabada, y es el asistente quien las traslada a una redacción solemne, correcta y perfectamente estructurada. El resultado es un borrador que, naturalmente, deberá revisarse y quizá modificarse en parte; pero el camino entre el análisis probatorio y jurídico ya hecho por el juez y la resolución redactada se acorta de manera notable. La ganancia no es de criterio —que sigue intacto en el juez—, sino de productividad y de tiempo.

Dos garantías esenciales atraviesan todo el método. La primera es la anonimización: hoy por hoy, todo el material —demandas, contestaciones, prueba y sentencias— debe subirse previamente anonimizado, extremo que conviene confirmar al inicio de cada sesión. No es una formalidad, sino la traducción práctica del límite de protección de datos y de la exigencia de que el expediente no salga del perímetro. Ahora bien, conviene hacer una precisión de futuro inmediato: esta exigencia es la regla prudente mientras persista la indefinición contractual, pero no tiene por qué ser definitiva. El día en que el Ministerio de Justicia o el Consejo General del Poder Judicial cierren un contrato en el que quede todo claro —la confidencialidad de los datos, la imposibilidad de emplearlos para enriquecer el modelo y la garantía de que no salen de un entorno cerrado y, además, radicado en la Unión Europea—, podrá empezarse a trabajar con total seguridad sin necesidad de anonimizar, pues se estará operando en un entorno verdaderamente privado; y ello redundará en una mayor celeridad en el uso de la inteligencia artificial. Es, precisamente, el escenario hacia el que apuntan los Copilot facilitados por el Ministerio. La segunda garantía son las reglas de rigor frente a la invención: el asistente no puede inventar jurisprudencia, números de sentencia, fechas ni preceptos; ante la duda debe advertirlo en lugar de fabricar la referencia; cuando el juez cite un precepto sin transcribirlo, ha de verificar su tenor y vigencia y, si no puede hacerlo con certeza, declararlo; y toda afirmación fáctica debe poder reconducirse al documento del que procede, con respeto a la congruencia, resolviendo todas las pretensiones y solo ellas.

Así entendido, el método preserva precisamente aquello que el artículo 117 CE reclama: que la formación de la convicción, la valoración de la prueba, la interpretación del Derecho y la responsabilidad del fallo permanezcan en el juez. La inteligencia artificial amplía la memoria, acelera la búsqueda, ordena el expediente y mejora la calidad formal de la resolución; no forma el juicio.

V. OTROS MODELOS GENERATIVOS SOBRE FUENTE PÚBLICA: DEL CENDOJ A LA CARPETA DE JURISPRUDENCIA

El marco descrito no agota el panorama. Junto al Copilot institucional existen otros grandes modelos de lenguaje de uso general —ChatGPT, Claude, Gemini o NotebookLM, entre los más difundidos—, a los que conviene aplicar una cautela de partida. Todos ellos, y tam-

bién el propio Copilot, pertenecen a lo que se conoce como inteligencias artificiales de «caja negra»: se desconocen los millones de datos con los que han sido entrenadas, de modo que su funcionamiento interno no es plenamente transparente ni verificable. Esa opacidad obliga a un cierto control en su empleo y explica que la Instrucción no las contemple entre los sistemas facilitados; no pueden, por ello, utilizarse sobre los datos protegidos del expediente. Pero de esa exclusión no se sigue que carezcan de utilidad legítima para el juez: la tienen, y notable, siempre que su empleo se circunscriba a materiales que no comprometan la confidencialidad del procedimiento concreto.



“Pero de esa exclusión no se sigue que carezcan de utilidad legítima para el juez: la tienen, y notable, siempre que su empleo se circunscriba a materiales que no comprometan la confidencialidad del procedimiento concreto”

Y aquí conviene precisar el alcance exacto de la restricción, porque suele entenderse de manera más amplia de lo que en rigor procede. La limitación que impone la Instrucción 2/2026 respecto de estos modelos no facilitados se ciñe a un terreno muy concreto: la subida de documentos no anonimizados y, señaladamente, de los escritos o las pruebas de parte del procedimiento de que se trate. Ahí —y solo ahí— opera la prohibición. Fuera de ese terreno, nada impide plantear a estos modelos cuestiones en abstracto, esto es, preguntas análogas a las que cualquier jurista formularía a un repertorio o a un compañero, con la ventaja añadida de que las versiones contratadas y de pago suelen ser sensiblemente más eficientes y ofrecer respuestas de mayor calidad.

Los ejemplos son cotidianos y, a mi juicio, perfectamente legítimos. Lo que no está autorizado es servirse de estos modelos para redactar una sentencia o una resolución concreta; pero sí cabe pedirles que localicen sentencias con un determinado contenido, que analicen trabajos doctrinales sobre una concreta cuestión o que precisen cuál es la jurisprudencia vigente en una materia, exigiéndoles siempre que indiquen la fuente. Cuando, además de la cita, facilitan el enlace o la referencia precisa, puede comprobarse sin esfuerzo la existencia real de la resolución y su contenido exacto —y otro tanto con los trabajos doctrinales—, lo que convierte la exigencia de fuente en el antídoto natural frente a la alucinación: solo se incorpora aquello que previamente se ha cotejado.

Y sobre todo ese material cabe, después, formular preguntas. No las del caso concreto, que en principio parece vedado, pero sí cuestiones planteadas en un plano más genérico, sin aportar dato alguno —ni del demandante ni del demandado—, e incluso supuestos de laboratorio análogos al que se tiene entre manos. La utilidad es enorme: el modelo devuelve respuestas que el juez

puede cotejar con sus propias fuentes y que, en no pocas ocasiones, resultan verdaderamente sugerentes. Es, de nuevo, conocimiento que se busca y se controla, no decisión que se delega. Esta utilización, en mi opinión, es innegociable: no vulnera ninguno de los principios de la Instrucción 2/2026 y resulta de enorme provecho, sin perjuicio de que después uno reescriba el resultado o se limite a aprovechar la idea al resolver el pleito. Limitarla sería confundir la prudencia debida con una renuncia carente de fundamento.

Un campo especialmente fértil es el tratamiento de la jurisprudencia ya publicada y anonimizada en el Centro de Documentación Judicial (CENDOJ). Sobre ese material —público por definición— estos modelos permiten elaborar resúmenes estructurados mediante prompts diseñados al efecto, ordenados en torno a los hechos no controvertidos y controvertidos, los fundamentos de derecho, el fallo, la doctrina jurisprudencial que se fija y las cuestiones de interés que la resolución suscita. Con ello cabe construir verdaderas carpetas de jurisprudencia, sistematizadas y de consulta inmediata, que multiplican el rendimiento del estudio doctrinal sin tocar dato protegido alguno.

De este modo, quienes van incorporando estas herramientas a su método de trabajo obtienen una doble ganancia de tiempo: por un lado, una mayor rapidez en la localización y el contraste de las fuentes; por otro —ya en el entorno facilitado por el Ministerio de Justicia— una redacción más ágil de las resoluciones y de los resúmenes de los escritos de parte. La eficiencia, conviene insistir, no es aquí un fin en sí mismo, sino el resultado de un trabajo previo que el juez sigue dirigiendo.

VI. DEL «PROMPT» AL «LOOP»: LA FORMACIÓN COMO VERDADERA CLAVE

Hay, por último, una transformación de fondo que conviene nombrar, porque altera la naturaleza misma de la herramienta: se ha pasado del «prompt» al «loop». La inteligencia artificial generativa ya no se limita a devolver una respuesta a una orden aislada; ha alcanzado un nivel en el que es posible sostener con ella una relación interactiva, un diálogo iterativo en el que cada respuesta se afina, se corrige y se reorienta a partir de la anterior. La herramienta ha dejado de ser un buscador para convertirse en un interlocutor técnico.

Y aquí reside la verdadera clave, que debe ser comprendida: en ese diálogo, los grandes beneficiados son quienes están especialmente formados en la materia. No es lo mismo la inteligencia artificial en manos de quien es lego que en las de quien posee altos conocimientos, como se presupone de jueces y magistrados. Una búsqueda de jurisprudencia, de conceptos o de doctrina realizada por un ciudadano de a pie no es equiparable a la que ejecuta un magistrado que sabe qué fuentes subir, dónde se encuentran las resoluciones o la doctrina pertinentes y cómo plantear la pregunta con la precisión técnica adecuada. El resultado de este segundo no procede del desconocimiento: es un resultado buscado, dirigido y depurado por el propio jurista que dialoga con la máquina.

De ahí que la formación —jurídica y, cada vez más, también en el manejo crítico de estas herramientas— sea hoy el factor decisivo. La inteligencia artificial no nivela hacia abajo ni suple el saber: amplifica el que ya se posee. Por eso su incorporación responsable a la función jurisdiccional no es un problema meramente tecnológico, sino, ante todo, de competencia profesional: cuanto mayor es el dominio del juez sobre la materia, mayor es también su

capacidad para extraer de la herramienta un rendimiento fiable y para advertir cuándo le ofrece un espejismo. La máquina sirve mejor a quien sabe lo que busca.

VII. CONCLUSIÓN: INTELIGENCIA ARTIFICIAL SIN ABDICACIÓN JURISDICCIONAL

La inteligencia artificial será constitucionalmente admisible como instrumento de apoyo; será sospechosa como arquitectura de condicionamiento; y será incompatible con la jurisdicción cuando opere como sustituto material del juicio humano. La legitimidad de una sentencia no se mide por su forma documental, sino por la realidad del juicio que la precede. La tarea que se ofrece al jurista no es optar entre tecnofobia y tecnoutopía —ambas, formas de eludir la responsabilidad—, sino un trabajo paciente de delimitación, en el que la implicación crítica de los jueces destinatarios de las herramientas es ya parte ineludible del ejercicio responsable de la función.

Y si algo resulta claro, a estas alturas, es que lo que se necesita son talleres: muchos, continuos y repartidos por todo el territorio nacional, en los que quienes han ido adquiriendo mayor formación en la materia puedan ir formando a quienes han quedado atrás. La inteligencia artificial generativa como ayuda en la elaboración de resoluciones no es un adorno ni una moda pasajera: es una herramienta fundamental y un avance de primer orden que, bien empleado, se traduce en un aumento real de la productividad y de la calidad de las resoluciones.

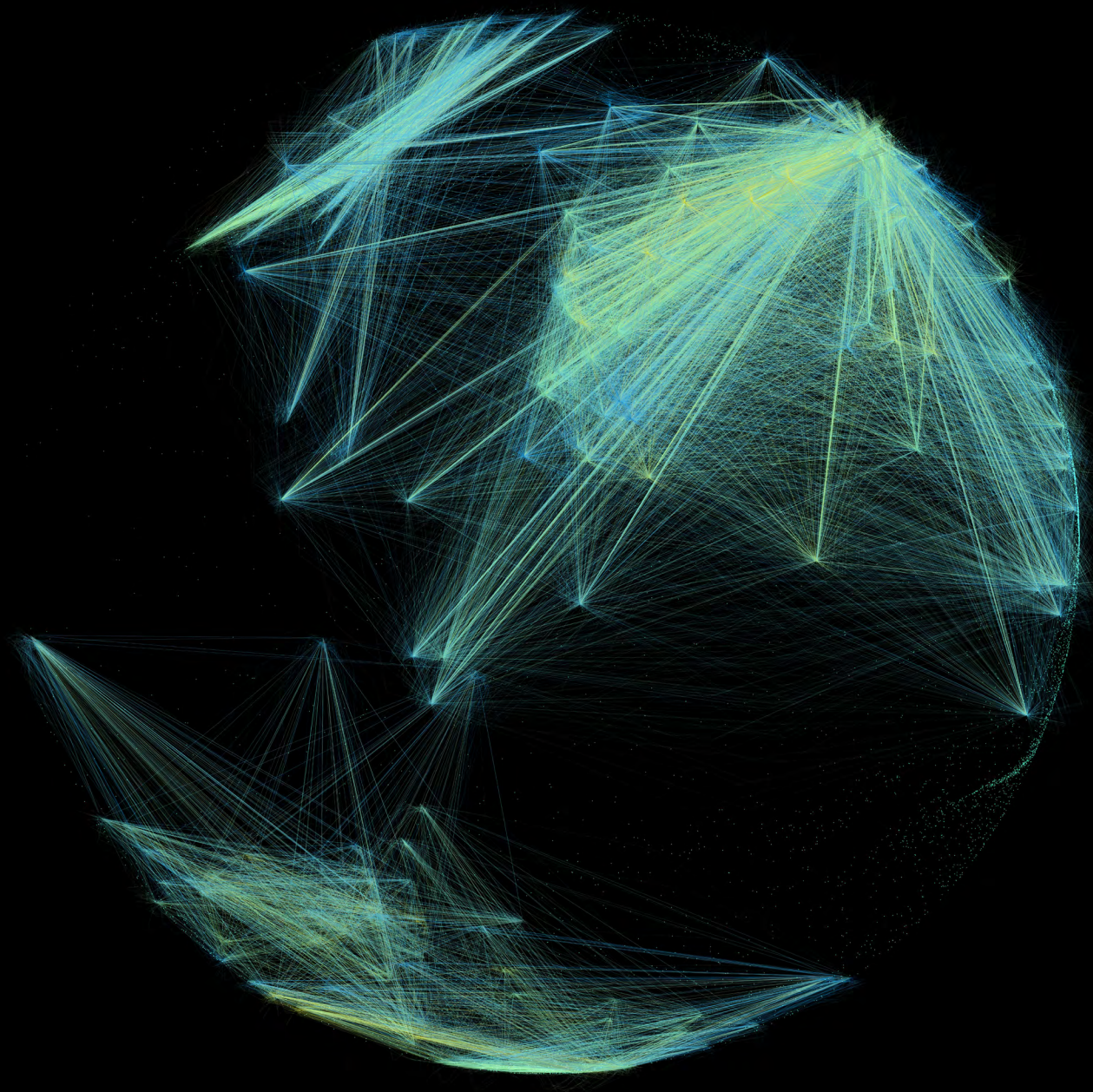
Esa misma potencia explica por qué la formación resulta imprescindible. No es algo usual entre jueces y magistrados, pero no cabe descartar que quien no esté bien instruido haga un uso indebido, no permitido o, sencillamente, inconstitucional de la herra-

mienta. El riesgo más insidioso es dejarse llevar: la IA puede producir resoluciones aparentemente motivadas, formalmente impecables, que colman de satisfacción a quien las emplea precisamente porque lo parecen, aunque sean equivocadas. De ahí a una tendencia al acomodamiento del juzgador —un sesgo de complacencia— media un solo paso. Y el juez no puede olvidar nunca cuál es el núcleo jurisdiccional que el artículo 117 CE no le permite ceder en ningún caso a una máquina.

La conclusión, por tanto, es nítida: el Consejo General del Poder Judicial debería tomar la cabeza. Más allá de algún curso de formación descentralizada o de alguna actividad de ámbito estatal, lo que procede es establecer talleres continuos, exactamente como en su día se hizo para el manejo del propio CENDOJ o de las herramientas de tramitación procesal. No puede permitirse que un solo juez o magistrado quede sin esa formación, porque de ella depende, en buena medida, el incremento de productividad y de calidad que la buena utilización de la herramienta promete.

La fórmula puede ser sencilla. La IA puede ayudar al juez a conocer mejor; no puede decidir por él. Puede asistir a la jurisdicción; no puede convertirse en jurisdicción invisible. El artículo 117 CE no garantiza solo que la sentencia sea firmada por un juez: exige que haya sido realmente juzgada por un juez.





La independencia judicial en la era de la Inteligencia Artificial

Sesgos, opacidad y delegación
cognitiva en el oficio de juzgar

Sandra María Piñeiro Vilas

Magistrada Presidenta de la Audiencia Provincial de Lugo

1. ¿QUÉ PUEDE HACER LA IA? ¿QUÉ NO DEBEMOS DEJAR DE HACER NOSOTROS?

La conversación sobre inteligencia artificial y Justicia suele comenzar con una enumeración de utilidades: localizar jurisprudencia, resumir escritos, ordenar expedientes, traducir documentos, transcribir vistas, identificar cuestiones controvertidas o generar un primer borrador. Esa aproximación es comprensible. La carrera judicial trabaja sometida a una presión creciente, con expedientes extensos, información fragmentada y exigencias temporales que no siempre son compatibles con la serenidad que reclama el juicio. Una herramienta capaz de reducir tareas mecánicas y recuperar tiempo para la decisión merece ser tomada en serio.

Pero la pregunta decisiva no es cuánto trabajo puede ahorrar. Es qué operaciones intelectuales y procesales no pueden ser externalizadas sin alterar la naturaleza de la jurisdicción. Juzgar no consiste únicamente en producir un texto jurídicamente motivado resolviendo la cuestión litigiosa. Supone fijar los hechos tras escuchar a las partes, valorar la prueba conforme a reglas jurídicas, identificar la norma aplicable, justificar la elección entre interpretaciones posibles y asumir públicamente la responsabilidad por la decisión. La resolución es el resultado visible de ese proceso, pero no lo agota. El núcleo jurisdiccional reside en la formación libre, individualizada y responsable del juicio.

La Instrucción 2/2026 del Consejo General del Poder Judicial lo formula con precisión al recordar que verificar los hechos, subsumirlos en la norma, seleccionar e interpretar el Derecho y emitir el pronunciamiento son operaciones que deben realizar los miembros de la judicatura desplegando las capacidades intelectivas y volitivas que conforman el núcleo de la función de juzgar¹. La IA es una ayuda; no un sujeto de potestad jurisdiccional. Esta distinción no es retórica. Permite entender que la frontera entre asistencia y sustitución no depende del nombre comercial de la herramienta ni de que exista formalmente una persona que pulse «aceptar». Depende del lugar que la herramienta ocupa en la cadena real de decisión.

Un sistema que clasifica documentos puede ser aparentemente auxiliar y, sin embargo, excluir de la atención judicial una prueba decisiva. Un resumidor puede omitir la excepción que altera el resultado. Un buscador jurídico puede ordenar precedentes según criterios comerciales o estadísticos no visibles. Un generador de texto puede proponer una estructura argumental que ancle al juez en una solución antes de completar el estudio. Un sistema predictivo puede transformar una probabilidad grupal en un juicio implícito sobre una persona concreta. En todos estos casos la influencia puede ser intensa, aunque formalmente la decisión final siga atribuida a un juez.

Se trata de evitar una modernización aparente que mejore la velocidad a costa de degradar la calidad constitucional del juicio. En Justicia, la eficiencia no es un valor aislado. Solo es legítima cuando sirve a una decisión más accesible, más comprensible y más respetuosa con los derechos.

1. Acuerdo de 28 de enero de 2026, del Pleno del Consejo General del Poder Judicial, por el que se aprueba la Instrucción 2/2026, sobre la utilización de sistemas de inteligencia artificial en el ejercicio de la actividad jurisdiccional, apartados séptimo.1, 2 y 4, y octavo.c), BOE núm. 27, de 30 de enero de 2026, BOE-A-2026-2205

1.1. De la independencia orgánica a la independencia cognitiva

La independencia judicial ha sido construida históricamente como una garantía frente a interferencias identificables: el poder político que pretende orientar el fallo, la jerarquía que condiciona al juez, el interés económico que busca capturar la decisión o la presión mediática que sustituye el expediente por el clima de opinión. Esa concepción sigue siendo imprescindible, pero se revela hoy como insuficiente. La mediación tecnológica introduce formas de influencia menos visibles. No ordenan decidir de una determinada manera: seleccionan lo que merece ser visto, jerarquizan las fuentes, proponen la primera hipótesis, convierten una correlación en un indicio o presentan una respuesta con una seguridad lingüística que invita a aceptarla.

Puede hablarse, por ello, de independencia cognitiva: la capacidad real del juez o de la jueza para formar su juicio a partir del conocimiento suficiente del asunto, sin que la arquitectura de una herramienta determine subrepticamente qué hechos, normas y argumentos entran en su campo de atención. No se trata de defender una imposible pureza mental. Todo juicio se forma mediante lenguajes, instituciones, precedentes y prácticas compartidas. La cuestión es otra: que esas mediaciones puedan ser conocidas, discutidas y corregidas, y que ninguna de ellas se convierta en una autoridad invisible.

Esta idea prolonga, en el terreno tecnológico, la noción clásica de independencia.

El Tribunal de Justicia, en la sentencia de 27 de febrero de 2018, *Associação Sindical dos Juízes Portugueses*, C-64/16, EU:C:2018:117, ap. 42, ha señalado que «la garantía de independencia, que es inherente a la misión de juzgar, obliga, en el ámbito de los Estados miembros, en lo que respecta a los jueces y tribunales nacionales». Además el TJUE, ap. 44, refiriéndose a los asuntos Wilson² y Margarit Panicello³, indicó que «la noción de independencia supone, entre otras cosas, que el órgano en cuestión ejerza sus funciones jurisdiccionales con plena autonomía, sin estar sometido a ningún vínculo jerárquico o de subordinación respecto a terceros y sin recibir órdenes ni instrucciones de ningún tipo, cualquiera que sea su procedencia, de tal modo que quede protegido de injerencias o presiones externas que puedan hacer peligrar la independencia de sus miembros a la hora de juzgar o que puedan influir en sus decisiones». El TJUE viene a establecer así una obligación para los Estados miembros de garantizar y respetar la independencia de sus tribunales nacionales, basándose en el artículo 19.1 TUE interpretado a la luz de los artículos 2 y 4.3 TUE. Lo que determina, en suma, la obligación de los Estados miembros de asegurarse de que sus órganos jurisdiccionales ofrezcan una tutela judicial efectiva a fin de garantizar el cumplimiento del Derecho de la Unión es inherente a un Estado de Derecho⁴.

El Tribunal Europeo de Derechos Humanos ha subrayado, a su vez, que la independencia no es un privilegio de quienes juzgan, sino una garantía de las personas justiciables y del Estado de Derecho.⁵ Las

2. STJUE de 19 de septiembre de 2006, *Wilson*, C-506/04, EU:C:2006:587, apartado 49.

3. STJUE de 16 de febrero de 2017, *Margarit Panicello*, C-503/15, EU:C:2017:126.

4. Dictamen del Servicio Jurídico del Consejo n° 13593/18, 25 de octubre de 2018, sobre la compatibilidad con los Tratados de la Unión de la propuesta de Reglamento del Parlamento Europeo y del Consejo sobre la protección del presupuesto de la Unión en caso de deficiencias generalizadas del Estado de Derecho en los Estados miembros, apdo. 29.

5. STEDH (Gran Sala), *Guðmundur Andri Ástráðsson c. Islandia*, demanda 26374/18, 1 de diciembre de 2020.

amenazas algorítmicas no encajan literalmente en el modelo de la orden externa. Sin embargo, una dependencia técnica que haga prácticamente imposible comprender, contradecir o abandonar la recomendación puede erosionar la misma facultad de decidir autónomamente que aquellas doctrinas protegen.

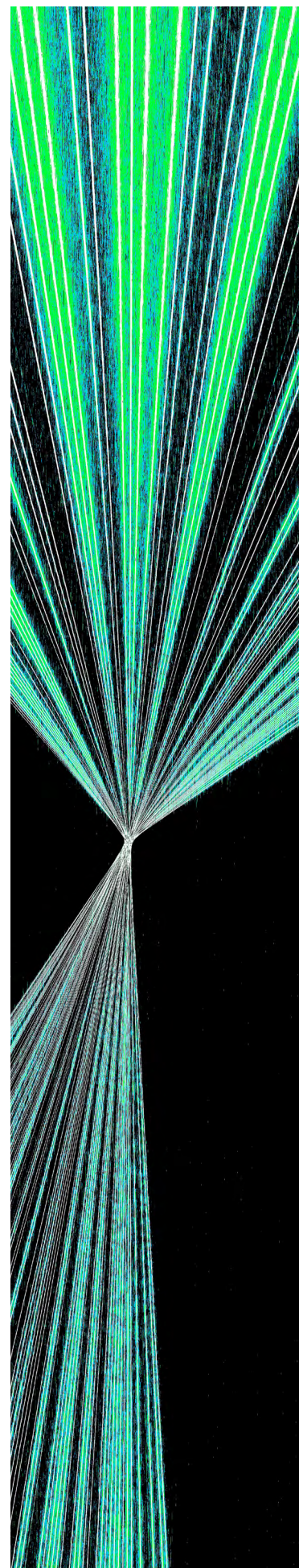
La independencia cognitiva tiene al menos cuatro dimensiones.

1. La primera es epistémica: conservar acceso directo y suficiente a las fuentes del caso. Conocer un expediente no equivale a conocer el resumen producido por un sistema. Resumir supone elegir relevancias, comprimir ambigüedades y omitir detalles; es ya una operación interpretativa.
2. La segunda es deliberativa: disponer de tiempo y condiciones para elaborar y contrastar hipótesis, en lugar de limitarse a confirmar la primera respuesta sugerida.
3. La tercera es decisoria: poder apartarse de la recomendación sin fricciones técnicas, organizativas o estadísticas que conviertan la discrepancia en una anomalía.
4. La cuarta es expresiva: ofrecer razones propias, comprensibles y conectadas con las alegaciones y pruebas del proceso.

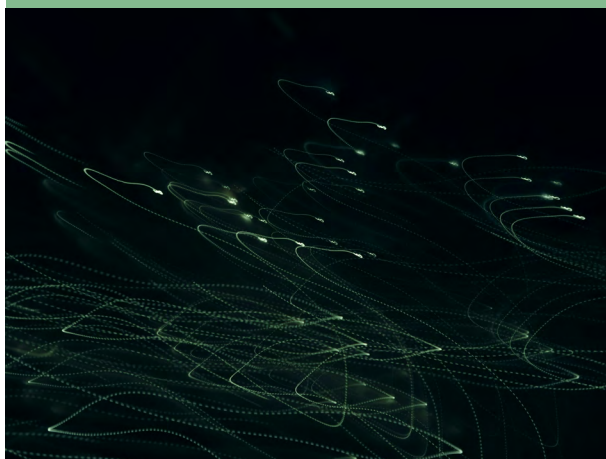
Esta concepción se inscribe en la denominada perspectiva humanocéntrica, frente a la tentación de abordar la IA como una mera carrera de capacidades.⁶ Esa convergencia entre ontología, ética y Derecho resulta especialmente fértil para la jurisdicción: antes de preguntar qué puede hacer una herramienta, obliga a decidir qué idea de persona, juicio y responsabilidad queremos preservar.

Aquí resulta sugerente —aunque deba manejarse como analogía y no como traslación literal— el problema planteado por Jürgen Habermas en *El futuro de la naturaleza humana*. Su objeto inmediato era la intervención biotecnológica sobre la dotación genética de quienes no pueden consentirla. Habermas temía que la persona programada

6. A. M. Marcos del Cano (coord.), *Inteligencia artificial. Desafíos éticos, sociales y jurídicos*, Dykinson, Madrid, 2026; A. M. Marcos del Cano, «¿Quo vadis inteligencia artificial?: cuestiones ontológicas y éticas», pp. 37-64; A. Tomás López, «Gobernanza humanista de la inteligencia artificial: fundamentos filosóficos, algorética y marcos regulatorios», pp. 65-90; M. de Asís Pulido, «Júpiter, Hércules y Hermes ante la inteligencia artificial jurídica», pp. 111-136.



“La IA judicial no modifica el genoma, pero puede producir otra forma de heterodeterminación: encerrar a alguien en la identidad estadística que otros datos y decisiones pasadas han construido para él”



dejase de reconocerse enteramente como autora indivisa de su propia vida y que se quebrase la simetría moral entre sujetos libres e iguales.⁷ La IA judicial no modifica el genoma, pero puede producir otra forma de heterodeterminación: encerrar a alguien en la identidad estadística que otros datos y decisiones pasadas han construido para él.

La analogía muestra un límite ético y jurídico esencial. La persona juzgada no puede ser tratada como un objeto cuyo comportamiento pueda predecirse y optimizarse, sino como alguien capaz de explicar su

conducta, contradecir las conclusiones del sistema y aportar razones. Un perfil de riesgo se construye a partir del pasado de un grupo; el proceso judicial, en cambio, debe atender a la historia singular de cada persona, a las pruebas y a la posibilidad de cambio. La dignidad exige no solo que la decisión final sea adoptada por un ser humano, sino también que, durante todo el procedimiento, la persona afectada sea escuchada y tratada como sujeto de derechos y razones.

No basta, por tanto, con proclamar que *el ser humano conserva la última palabra*. Esa fórmula puede ocultar un control puramente ceremonial. Si la persona recibe un resultado cuando ya no tiene tiempo para reconstruirlo; si ignora los datos, las variables o la tasa de error; si la interfaz presenta la recomendación como respuesta objetiva; si apartarse exige una motivación adicional mientras aceptarla es automático; o si la organización mide su productividad por el grado de adhesión al sistema, la última palabra puede seguir siendo humana y la decisión haber dejado de serlo en sentido material. El control efectivo requiere competencia, información, tiempo, autoridad y posibilidad de discrepancia.

1.2. Juzgar no es predecir: la dimensión práctica del razonamiento jurídico

La IA contemporánea invita a confundir dos operaciones distintas: anticipar lo que probablemente decidirá un tribunal y determinar lo que jurídicamente debe decidirse en un caso. La primera es descriptiva y estadística; la segunda es normativa. Una frecuencia de decisiones pasadas puede informar sobre una práctica, pero no con-

7. J. Habermas, *El futuro de la naturaleza humana. ¿Hacia una eugenesia liberal?*, trad. R. S. Carbó, Paidós, Barcelona, 2002, especialmente pp. 29-100. La aplicación al perfilado algorítmico que aquí se propone es una analogía doctrinal propia.

vierte esa práctica en la respuesta correcta. El dato de que una determinada alegación fue rechazada en un porcentaje elevado de casos no explica si debe ser rechazada en el asunto actual, ni permite saber si la regularidad histórica expresa un precedente valioso o una injusticia reiterada.

La diferencia remite al carácter práctico del razonamiento jurídico. Para Robert Alexy, la decisión jurídica formula una pretensión de corrección y debe poder justificarse mediante razones; no es la simple constatación causal de lo que suele suceder.⁸ En Dworkin, el juez no obtiene la solución aplicando mecánicamente reglas, sino interpretando la práctica jurídica a la luz de los principios que mejor la justifican.⁹ Y Taruffo recordó que la decisión sobre los hechos no puede reducirse a intuición ni a cálculo: requiere una justificación racional e intersubjetivamente controlable de la valoración probatoria.¹⁰

Estas concepciones discrepan en aspectos importantes, pero convergen en algo decisivo para el presente análisis: *juzgar implica dar razones y asumir la responsabilidad de su suficiencia*. Un modelo puede reconocer patrones, recuperar textos y generar argumentos plausibles. No comparece, sin embargo, ante las partes, no participa en la práctica institucional como sujeto responsable y no responde por la pretensión de corrección de lo que produce. La fluidez lingüística de una IA generativa no debe confundirse con comprensión jurídica ni con compromiso institucional.

El juez Hércules de Dworkin ofrece aquí una metáfora fértil si se utiliza con cautela. Su capacidad sobrehumana para reconstruir la integridad del Derecho puede recordar a la promesa contemporánea de procesar toda la legislación y jurisprudencia disponibles. Pero Hércules no es una máquina de coincidencias: es un artificio filosófico para mostrar el peso de la interpretación y los principios. Una IA entrenada con decisiones pasadas podría imitar el estilo de Hércules y comportarse, en realidad, como Júpiter: un aplicador rígido de regularidades presentadas como reglas. Frente a ambos, la figura de Hermes —el juez de la red, de la circulación de sentidos y de la mediación— recuerda que el Derecho se construye comunicativamente y que ninguna respuesta puede separarse del procedimiento que la legitima.¹¹

1.3. La infraestructura también gobierna

La independencia cognitiva no es solo una virtud individual. Sería injusto exigir al juez que resista por sí solo una arquitectura diseñada para inducir adhesión. La adquisición, configuración y despliegue de sistemas de IA son decisiones de gobierno judicial. Determinan qué proveedores acceden al ecosistema, qué fuentes se indexan, qué usos se autorizan, qué registros se conservan, quién puede auditar, cómo se comunica el resultado y qué formación reciben los usuarios. La infraestructura distribuye poder.

Lessig popularizó la idea de que el código regula¹²; Pasquale describió la sociedad de la

8. R. Alexy, *Teoría de la argumentación jurídica*, cit.; sobre la pretensión de corrección como elemento del Derecho, véase también *El concepto y la validez del Derecho*, Gedisa, Barcelona, 1994.

9. R. Dworkin, *Law's Empire*, Harvard University Press, 1986, especialmente la concepción del Derecho como integridad.

10. M. Taruffo, *La prueba de los hechos*, Trotta, 2002.

11. F. Ost, *Júpiter, Hércules, Hermes: tres modelos de juez*, Academia, Revista sobre enseñanza del Derecho, año 4, nº 8, 2006, pág. 101-130.

12. L. Lessig, *El código y otras leyes del ciberespacio*, Taurus, Madrid, 2001.

caja negra, en la que actores públicos y privados clasifican a las personas mediante lógicas inaccesibles¹³; Hildebrandt ha insistido en que la arquitectura computacional no se limita a ejecutar normas, sino que reconfigura las posibilidades de acción y conocimiento.¹⁴ En el ámbito judicial, esta observación adquiere intensidad constitucional. Quien diseña el buscador decide, mediante criterios de relevancia, qué precedentes aparecen primero. Quien fija el umbral de una alerta distribuye falsos positivos y falsos negativos. Quien configura la interfaz decide si la incertidumbre se ve o queda escondida tras un color. Quien controla las actualizaciones puede modificar silenciosamente la herramienta que interviene en miles de asuntos.

De ahí que la independencia tecnológica requiera también capacidad pública: conocimiento interno, contratación responsable, interoperabilidad, auditoría independiente, control de versiones y reversibilidad. La dependencia de un proveedor no es un inconveniente comercial menor cuando afecta a la potestad jurisdiccional. Puede convertirse en una dependencia institucional si el sistema no es explicable, trasladable o sustituible sin perder acceso a los datos, registros o criterios con los que se adoptaron decisiones.

La tesis puede formularse de manera sencilla: la autonomía del juez depende también de la autonomía de la institución frente a la tecnología que utiliza. La garantía no se satisface prohibiendo toda herramienta externa, pero exige que ninguna organización pública entregue a un proveedor privado la capacidad de definir de manera opaca las condiciones cognitivas del juicio.

2. UN MAPA DE RIESGOS: LA IA JUDICIAL COMO SISTEMA SOCIOTÉCNICO

El error más frecuente consiste en tratar el algoritmo como un objeto aislado cuya bondad puede medirse exclusivamente por su porcentaje de acierto. La decisión asistida por IA es un sistema sociotécnico: incluye los datos históricos, las categorías con las que fueron etiquetados, el modelo, la interfaz, las instrucciones, el modo en que se presenta el resultado, la formación del usuario, las presiones organizativas y la posibilidad real de apartarse de la recomendación. Un modelo técnicamente correcto puede producir resultados injustos por la forma en que se despliega; uno menos preciso puede ser más respetuoso con las garantías si sus límites son visibles y su uso está estrictamente acotado.

Conviene distinguir cuatro niveles de riesgo.

1. El primero es el riesgo de entrada: datos incompletos, no representativos, obtenidos en contextos discriminatorios o contruidos mediante variables que no miden lo que dicen medir.
2. El segundo es el riesgo de modelo: elección de objetivos, ponderaciones, umbrales y métricas de error que incorporan decisiones normativas.

13. F. Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information*, Harvard University Press, Cambridge (Massachusetts), 2015, <https://doi.org/10.4159/harvard.9780674736061>.

14. Mireille Hildebrandt, *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology*, Edward Elgar Publishing, Cheltenham–Northampton, 2015, <https://doi.org/10.4337/9781849808774>.

3. El tercero es el riesgo de interacción: automatización complaciente, anclaje, sobreconfianza o uso de la herramienta fuera del contexto para el que fue validada.
4. El cuarto es el riesgo procesal e institucional: imposibilidad de contradicción, desigualdad entre las partes, ausencia de registros, dependencia del proveedor, actualización silenciosa del sistema o dificultad para atribuir responsabilidades.

Esta clasificación ayuda a comprender que eliminar el sesgo o hacer explicable el algoritmo no son operaciones únicas. Exigen actuar durante todo el ciclo de vida. El Reglamento europeo de IA sigue precisamente esa lógica: gestión continua del riesgo, gobernanza de datos, documentación técnica, registros, transparencia, supervisión humana, evaluación de impacto y vigilancia posterior. La pregunta no es solo si el sistema funciona, sino para quién funciona, en qué condiciones, con qué tasa de error, quién soporta los falsos positivos, qué ocurre cuando cambia la población y qué remedio existe cuando el sistema falla.

Mapa sintético de riesgos y garantías

Riesgo	Garantía comprometida	Respuesta institucional
Sesgo	Igualdad, no discriminación e individualización	Gobernanza de datos, pruebas por subgrupos y auditoría continua
Opacidad	Contradicción, motivación y recurso	Transparencia funcional, explicación y trazabilidad
Delegación	Juez independiente y responsabilidad personal	Control humano real, crítico y con tiempo suficiente
Confidencialidad	Privacidad y secreto jurisdiccional	Sistemas autorizados, minimización y seguridad

La evaluación debe recaer sobre el sistema completo: datos, modelo, interfaz, usuario, organización y remedios.

3. PRIMER RIESGO: LOS SEGOS ALGORÍTMICOS

3.1. El sesgo no nace solamente en el código

La Instrucción 2/2026 define el sesgo algorítmico como una desviación sistemática de los resultados derivada de los datos, modelos o procesos utilizados, susceptible de producir efectos arbitrarios o discriminatorios. La definición es útil porque evita identificar el sesgo únicamente con una intención discriminatoria del programador. Un sistema puede perjudicar de forma sistemática a un grupo sin que nadie haya introducido expresamente una variable prohibida.

La taxonomía de Suresh y Guttag¹⁵ —sesgos históricos, de representación, de medición, de agregación, de aprendizaje, de evaluación y de despliegue— permite localizar el daño en todo el ciclo de vida, no solo en el código.

15. H. Suresh; J. V. Guttag, *A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle*, Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '21), ACM, 2021, art. 17, pp. 1-9, DOI 10.1145/3465416.3483305.

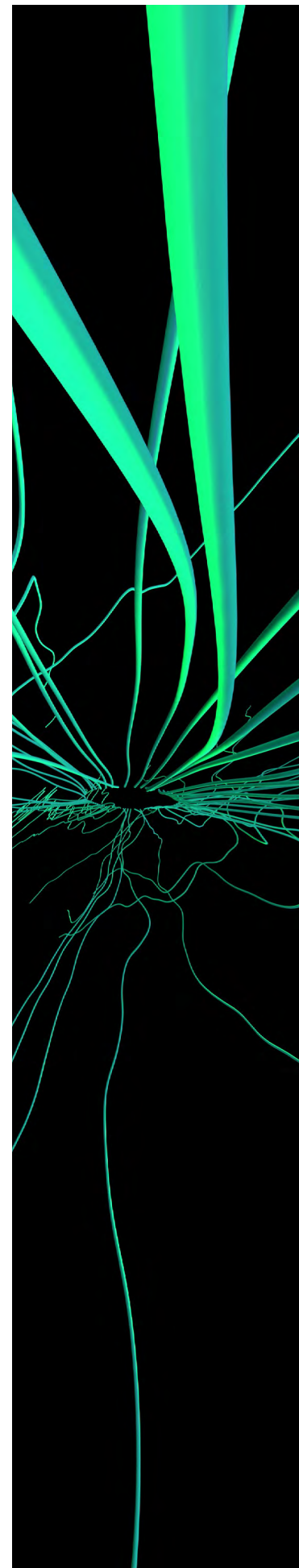
Existe *sesgo histórico* cuando los datos reproducen decisiones pasadas marcadas por desigualdades estructurales. Si un sistema aprende de detenciones, denuncias, concesiones de libertad provisional, sanciones disciplinarias o decisiones de protección adoptadas en contextos desiguales, puede convertir esas desigualdades en patrón predictivo. No aprende *la realidad* en abstracto; aprende los registros institucionales de una realidad observada a través de determinadas prácticas.

Existe *sesgo de representación* cuando ciertos colectivos aparecen de manera insuficiente o distorsionada en los datos. Un modelo entrenado principalmente con decisiones en castellano puede rendir peor en gallego, catalán o euskera. Una herramienta de análisis de testimonios puede estar optimizada para formas lingüísticas dominantes y tratar como anomalía la expresión de personas con discapacidad, menores, migrantes o hablantes no nativos. La ausencia de datos no genera neutralidad; genera invisibilidad.

Existe *sesgo de medición* cuando la variable utilizada es un sustituto imperfecto del fenómeno jurídico relevante. La reincidencia medida mediante nuevas detenciones no equivale a comisión de nuevos delitos. La «conflictividad» inferida de intervenciones policiales puede medir intensidad de vigilancia. El riesgo de incomparecencia puede confundirse con precariedad residencial o laboral. La credibilidad no puede deducirse de la fluidez verbal, el contacto visual o la expresión emocional sin introducir estereotipos culturales, de género o de discapacidad.

Existe *sesgo de agregación* cuando un único modelo se aplica a poblaciones heterogéneas. Una tasa global aceptable puede ocultar errores muy superiores en subgrupos pequeños. También existe cuando se ofrece una probabilidad construida para una población como si fuera una descripción individual. El paso de «personas con estas características presentan estadísticamente un riesgo mayor» a «esta persona es peligrosa» no es una inferencia automática; es una decisión normativa que exige individualización y cautela.

Existe *sesgo de despliegue* cuando la herramienta se utiliza para una finalidad distinta de aquella para la que fue diseñada o en un momento procesal diferente. Un sistema creado para orientar recursos de intervención social no debe convertirse, sin nueva validación, en fundamento de una medida cautelar o de una pena. La misma puntuación puede tener efectos radicalmente distintos según se use para ofrecer apoyo, restringir libertad o distribuir vigilancia.



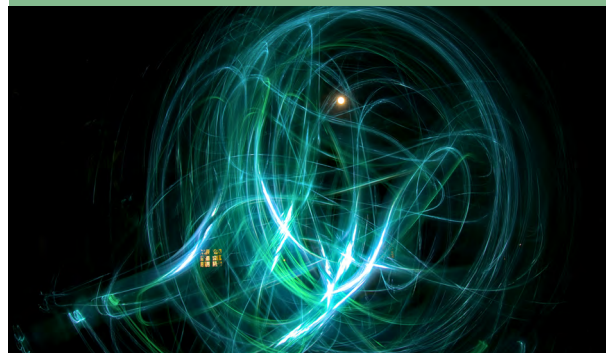
3.2. Las variables aparentemente neutras y el problema de los proxies

Excluir del modelo el sexo, el origen o la discapacidad no elimina necesariamente la discriminación. El código postal, el historial de empleo, el tipo de vivienda, la trayectoria educativa, el uso de determinados servicios públicos o la frecuencia de contactos policiales pueden operar como variables sustitutivas. En sociedades estructuradas por desigualdades, muchas características están correlacionadas con categorías protegidas.

El problema se agrava porque la correlación puede ser difícil de detectar en modelos complejos. Una variable aislada quizá parezca inocua, pero su combinación con otras reconstruye de manera aproximada la característica protegida. Además, retirar una variable puede empeorar la capacidad de comprobar el trato desigual. La respuesta jurídica no puede limitarse a ordenar que el modelo sea *ciego* a ciertas categorías; debe examinar los resultados, las tasas de error y el impacto sobre los grupos afectados.

En este punto resulta esencial diferenciar igualdad formal e igualdad material. Un sistema aplica la misma fórmula a todas las personas y, sin embargo, puede producir un impacto desproporcionado. La evaluación debe incluir, al menos, precisión por subgrupos, falsos positivos y falsos negativos, calibración, estabilidad temporal y posibilidad de reparación. En materia penal, migratoria, de menores o de violencia, no todos los errores tienen el mismo coste. Clasificar erróneamente a una persona como de alto riesgo puede restringir derechos; clasifi-

“Excluir del modelo el sexo, el origen o la discapacidad no elimina necesariamente la discriminación”



carla como de bajo riesgo puede afectar a la protección de terceros. Elegir el umbral es una decisión de política jurídica, no una cuestión meramente técnica.

3.3. No existe una única definición matemática de «equidad»

La investigación sobre justicia algorítmica ha demostrado que varias métricas de equidad pueden ser incompatibles cuando las tasas de base difieren entre grupos. Los trabajos de Kleinberg, Mullainathan y Raghavan¹⁶, y de Chouldechova¹⁷, muestran que no siempre es posible mantener simultáneamente calibración y paridad en las tasas de error. Esto tiene una consecuencia de enorme importancia: el proveedor no puede afirmar que el sistema es «justo» sin

16. Kleinberg, J.; Mullainathan, S.; Raghavan, M., *Inherent Trade-Offs in the Fair Determination of Risk Scores*, ITCS 2017, art. 43, [DOI 10.4230/LIPIcs.ITCS.2017.43](https://doi.org/10.4230/LIPIcs.ITCS.2017.43).

17. Chouldechova, A., *Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments*, *Big Data*, vol. 5, n.º 2, 2017, pp. 153-163, [DOI 10.1089/big.2016.0047](https://doi.org/10.1089/big.2016.0047)

explicar qué noción de justicia ha elegido y qué costes distribuye esa elección.

La decisión sobre qué error debe minimizarse no puede quedar oculta dentro del modelo. Si se prioriza reducir falsos negativos, aumentarán normalmente los falsos positivos. Si se busca la misma tasa de falsos positivos entre grupos, puede alterarse la calibración. En un proceso judicial, esa elección debe conectarse con los derechos en juego, la presunción de inocencia, la proporcionalidad y la carga de la prueba. La estadística puede describir el compromiso; no puede decidir por sí sola cuál es constitucionalmente aceptable.

Por eso las auditorías de sesgo no deben consistir en una cifra única. Han de documentar las métricas escogidas, justificar su pertinencia jurídica y mostrar las consecuencias distributivas. Es especialmente peligroso convertir la «precisión global» en certificado de legitimidad. Un sistema con un 90 % de acierto puede ser inaceptable si concentra el 10 % de errores en un colectivo vulnerable o si los errores recaen siempre en la misma dirección.

3.4. El precedente como fuente de conservadurismo algorítmico

Los sistemas jurídicos generativos y predictivos se alimentan de resoluciones previas. Esa base ofrece una ventaja evidente: permite localizar patrones, doctrina consolidada y líneas jurisprudenciales. Pero también introduce un riesgo menos visible: transformar el pasado en límite de lo jurídicamente imaginable.

Los datos judiciales reflejan qué asuntos llegaron a los tribunales, quién pudo litigar

y qué argumentos fueron formulados. Si la herramienta ordena por frecuencia, puede presentar como «normal» lo meramente mayoritario y relegar decisiones innovadoras, votos particulares o jurisprudencia supranacional.

Este fenómeno afecta especialmente a la interpretación evolutiva de los derechos fundamentales. La igualdad, la perspectiva de género, la protección de la infancia o los derechos de las personas con discapacidad han avanzado precisamente porque los tribunales cuestionaron categorías heredadas. Una IA entrenada para reproducir el centro estadístico de la jurisprudencia puede favorecer una suerte de inmovilismo probabilístico: la respuesta más frecuente aparece como la más correcta.

La garantía frente a este riesgo exige que las herramientas no solo recuperen precedentes similares, sino que permitan identificar excepciones, cambios normativos, votos discrepantes y líneas contradictorias. El juez debe poder conocer por qué determinados resultados aparecen primero y qué universo documental ha quedado fuera. Una búsqueda jurídicamente adecuada no es la que confirma con mayor rapidez la hipótesis inicial, sino la que facilita también encontrar razones para refutarla.

3.5. La interacción entre sesgo humano y sesgo algorítmico

La literatura empírica obliga a desconfiar de una idea intuitiva: añadir un algoritmo a una decisión humana no produce automáticamente una decisión más precisa ni más justa. Green y Chen proponen estudiar el *algoritmo en el circuito* (*algorithm in the loop*), es decir, la interacción real entre recomendación y decisor¹⁸. El diseño de la

18. B. Green y Y. Chen, *The Principles and Limits of Algorithm-in-the-Loop Decision Making*, *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, CSCW, art. 50, 2019, pp. 1-24, <https://doi.org/10.1145/3359152>.

interfaz, el orden de la información y la autoridad percibida del sistema pueden producir efectos distintos sobre grupos diferentes.

La automatización puede corregir intuiciones arbitrarias, pero también otorgarles una apariencia científica. Un juez puede utilizar la puntuación para confirmar una impresión previa y descartarla cuando la contradice. Puede confiar más en ella cuando la carga de trabajo es elevada o cuando el caso es rutinario. Puede sentirse protegido por el respaldo de la herramienta ante una decisión impopular. Estos usos selectivos no suelen aparecer en las métricas del proveedor.

El caso *State v. Loomis*, resuelto por el Tribunal Supremo de Wisconsin en 2016, constituye una referencia clásica. El tribunal permitió considerar la puntuación COMPAS con cautelas, pero insistió en que no podía ser determinante, en la necesidad de una sentencia individualizada y en la obligación de conocer las limitaciones del instrumento. La distinción entre «considerar» y «depender» del resultado es valiosa, aunque difícil de verificar en la práctica. Una recomendación puede influir de forma decisiva sin ser citada como fundamento autónomo.

El asunto merece una lectura más detenida. Eric Loomis alegó que el uso de COMPAS en la determinación de su pena vulneraba el *due process*: la lógica del producto estaba protegida como secreto comercial, la evaluación se construía mediante datos grupales y el instrumento incorporaba el género en su calibración. El Tribunal Supremo de Wisconsin no declaró inconstitucional su utilización, pero la rodeó de advertencias. COMPAS no podía emplearse para decidir el encarcelamiento ni para fijar la severidad de la pena; debía acompañarse de información sobre su carácter propietario, la imposibilidad de conocer la ponderación de factores, su fundamento en datos colectivos y los problemas de validación y sesgo.¹⁹

La sentencia es ambivalente. Su valor reside en reconocer que un instrumento predictivo no es una fuente neutra y que necesita límites explícitos. Su debilidad consiste en confiar en que una cláusula de advertencia neutralice la influencia cognitiva de una puntuación ya incorporada al informe. La psicología de la decisión muestra que conocer de forma abstracta una limitación no elimina el anclaje. Además, la prohibición de que el resultado sea «determinante» es difícil de controlar si no se documenta qué peso tuvo en el razonamiento real. Loomis no es, por ello, un certificado de legitimidad de los sistemas de riesgo; es una advertencia temprana sobre la tensión entre secreto empresarial, individualización y defensa.

3.6. Sesgos lingüísticos, de género e interseccionales

En España, la evaluación del sesgo debe atender a la pluralidad lingüística y territorial. Un modelo puede reconocer peor la terminología jurídica en lenguas cooficiales, perder matices en la traducción o recuperar menos precedentes. Si la herramienta se integra en el flujo ordinario de trabajo, esa menor calidad puede crear incentivos para abandonar el uso procesal de la lengua propia o para considerar menos fiables los documentos redactados en ella.

19. Wisconsin Supreme Court, *State v. Loomis*, 2016 WI 68, aps. 66-74 y 98-100. El recurso fue desestimado, pero el tribunal impuso límites de uso y una advertencia obligatoria sobre las limitaciones del instrumento.

La perspectiva de género tampoco puede limitarse a comprobar si el sexo aparece como variable. Los modelos de lenguaje absorben asociaciones presentes en enormes corpus: autoridad vinculada a lo masculino, cuidado a lo femenino, distinta valoración de la expresión emocional, normalización de estereotipos sobre víctimas o familias. En violencia sexual o de género, un sistema entrenado sobre resoluciones anteriores puede reproducir mitos y estereotipos que la jurisprudencia y la formación judicial tratan precisamente de superar.

Los sesgos se acumulan. Una mujer migrante, con discapacidad y que se expresa en una lengua no dominante no experimenta cada factor de forma separada. Las auditorías que solo comparan hombres y mujeres o grupos mayoritarios y minoritarios pueden ocultar daños interseccionales. Esto exige participación de los colectivos afectados, pruebas cualitativas y mecanismos de reclamación, no únicamente métricas agregadas.

La conclusión de este apartado puede formularse así: el sesgo no es un defecto accidental que se elimina limpiando datos. Es un riesgo estructural que exige decisiones normativas explícitas, auditoría continua y capacidad efectiva de suspender el sistema. La igualdad no puede delegarse en una métrica seleccionada por el proveedor.

4. SEGUNDO RIESGO: OPACIDAD ALGORÍTMICA Y DÉFICIT DE CONTRADICCIÓN

4.1. Tres formas de opacidad

Jenna Burrell²⁰ distinguió tres fuentes de opacidad en el aprendizaje automático. La primera es la opacidad *intencional*, derivada del secreto empresarial, la propiedad intelectual o la seguridad. La segunda es la opacidad *técnica*, producida por la falta de conocimientos necesarios para interpretar el sistema. La tercera es la opacidad *intrínseca*, propia de modelos cuya complejidad dificulta reconstruir de manera intuitiva por qué una combinación de parámetros generó un resultado.

En Justicia las tres se superponen. El proveedor puede negar acceso al modelo o a los datos de entrenamiento; los usuarios pueden carecer de formación para interpretar la documentación; y la explicación técnica puede no traducirse en una razón jurídicamente relevante. La transparencia, por tanto, no se satisface entregando miles de líneas de código. Una parte necesita saber qué información se utilizó, qué función cumplió el sistema, qué factores influyeron materialmente, cuáles son sus límites y cómo puede cuestionar el resultado.

La sentencia del Tribunal de Justicia de la Unión Europea de 27 de febrero de 2025, asunto C-203/22, *Dun & Bradstreet Austria*, ofrece una idea especialmente fecunda. La explicación debe permitir comprender el procedimiento y los principios aplicados, qué datos se utilizaron y de qué modo. El Tribunal advierte que la mera comunicación del algoritmo no constituye necesariamente una explicación concisa y comprensible. Incluso puede ser más útil una explicación contrafactual: indicar en qué medida una variación de los datos habría conducido a un resultado distinto.

20. Jenna Burrell, *How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms*, *Big Data & Society*, vol. 3, núm. 1, 2016, pp. 1-12, <https://doi.org/10.1177/2053951715622512>.

La transparencia jurídicamente relevante es funcional: debe permitir comprender, impugnar y revisar. Según el riesgo, exigirá código, documentación del modelo, datos de validación, registros, umbrales, versiones, instrucciones o fuentes recuperadas.

4.2. El precedente BOSCO: abrir el código, pero no detenerse en el código

La sentencia de la Sala Tercera del Tribunal Supremo de 11 de septiembre de 2025, dictada en el recurso 7878/2024, reconoció el acceso al código fuente de BOSCO, aplicación utilizada para comprobar requisitos del bono social eléctrico. El Tribunal conectó el derecho de acceso a la información pública con las exigencias democráticas y el Estado de Derecho, y afirmó una transparencia reforzada cuando una aplicación automatizada incide en derechos sociales de personas vulnerables.

BOSCO es un precedente decisivo porque rechaza que la invocación genérica de la propiedad intelectual o de la seguridad clausure el control. También porque muestra que el algoritmo no es una zona ajena al Derecho: si traduce requisitos legales a reglas operativas, debe poder comprobarse que esa traducción es correcta.

Sin embargo, no debe extraerse una conclusión simplista. En modelos de aprendizaje automático o generativos, el código fuente puede ser insuficiente. El comportamiento depende también de los datos, los pesos del modelo, la versión, las instrucciones del sistema, los documentos recuperados,

los parámetros de generación y las actualizaciones. Dos consultas idénticas pueden producir respuestas diferentes. Una auditoría útil debe reconstruir el entorno material de la decisión.

La transparencia tampoco exige publicar indiscriminadamente información que comprometa la seguridad o datos de terceros. Exige una ponderación concreta y soluciones graduadas: acceso parcial, peritaje bajo deber de confidencialidad, entrega al tribunal o a una autoridad independiente, documentación sustitutiva o pruebas controladas. El secreto comercial no desaparece, pero tampoco puede convertirse en inmunidad frente a la tutela judicial.

4.3. La opacidad propia de la IA generativa: fluidez, alucinación y falsa autoridad

Los modelos de lenguaje generan la secuencia verbal estadísticamente más plausible; no verifican por sí mismos la verdad jurídica. Su principal peligro no es que produzcan siempre errores, sino que pueden expresarlos con la misma seguridad y coherencia formal que una respuesta correcta. La fluidez reduce la percepción de riesgo.

Dahl, Magesh, Suzgun y Ho documentaron tasas elevadas de *alucinación jurídica* en modelos generalistas al responder preguntas verificables sobre jurisprudencia federal estadounidense²¹. Estudios posteriores sobre herramientas jurídicas con recuperación de fuentes mostraron que el uso de bases documentales reduce, pero no elimina, los errores²². Las cifras concretas no deben

21. Matthew Dahl, Varun Magesh, Mirac Suzgun y Daniel E. Ho, *Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models*, *Journal of Legal Analysis*, vol. 16, núm. 1, 2024, pp. 64-93, <https://doi.org/10.1093/jla/laae003>.

22. Varun Magesh, Faiz Surani, Matthew Dahl, Mirac Suzgun, Christopher D. Manning y Daniel E. Ho, «Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools», *Journal of Empirical Legal Studies*, vol. 22, núm. 2, 2025, <https://doi.org/10.1111/jels.12413>.

trasladarse automáticamente a modelos actuales ni al Derecho español; los sistemas evolucionan y los experimentos responden a tareas específicas. La conclusión estable es otra: ninguna arquitectura elimina la obligación de verificar cada cita, cada norma y cada afirmación fáctica.

El riesgo aumenta cuando se utiliza para resumir un expediente. Un resumen no es una reducción neutral: selecciona y jerarquiza. Puede omitir una alegación subsidiaria, confundir quién afirmó un hecho, transformar una incertidumbre en certeza o perder la secuencia temporal. Si el juez estudia únicamente el resumen, el error deja de ser auxiliar y pasa a configurar el objeto mismo de la decisión.

Por ello, las directrices de la CEPEJ de 2025 recomiendan herramientas especializadas, seguras y bajo control público; exigen verificación de citas, trazabilidad, evaluación previa de riesgos y responsabilidad judicial exclusiva²³. También advierten contra la utilización de sistemas generativos para cuestiones jurídicas novedosas, constitucionales o de especial complejidad normativa, precisamente porque la producción de una respuesta plausible puede ocultar la inexistencia de un patrón fiable.

4.4. Opacidad, motivación y recurso

Una resolución motivada no es aquella que contiene muchas razones, sino aquella que expone las razones reales y jurídicamente controlables de la decisión. Si el texto incorpora un argumento generado por IA que el juez no ha reconstruido y validado, puede existir una motivación solo aparente. La corrección estilística no compensa la ausencia de apropiación intelectual.

La opacidad afecta también al recurso. La parte debe poder identificar el itinerario racional para combatirlo. Si una puntuación, una clasificación o una selección algorítmica influyeron materialmente y no se conoce su intervención, el recurso se dirige contra una imagen incompleta de la decisión. La segunda instancia tampoco puede revisar de forma efectiva aquello que no figura en el expediente o cuyos registros no se conservaron.

23. Comisión Europea para la Eficacia de la Justicia (CEPEJ), *Guidelines on the Use of Generative Artificial Intelligence for Courts*, documento CEPEJ(2025)18Final, Consejo de Europa, Estrasburgo, 19 de diciembre de 2025



La exigencia de trazabilidad no significa incorporar a la resolución cada interacción irrelevante con una herramienta. Significa conservar la información necesaria cuando la IA haya influido materialmente: sistema y versión, finalidad, documentos de entrada, fuentes recuperadas, resultado relevante, intervención humana y controles realizados. Sin ese rastro, la afirmación de que «el juez decidió» puede ser jurídicamente cierta en sentido formal, pero imposible de verificar en sentido sustancial.

4.5. Igualdad de armas y asimetría tecnológica

La opacidad adquiere una dimensión procesal cuando una parte o el propio tribunal dispone de herramientas inaccesibles para la otra. No existe un derecho general a utilizar idéntica tecnología, pero sí a conocer y controvertir los elementos que influyen en la decisión. Si la Administración aporta una valoración algorítmica, la defensa debe poder acceder a información suficiente sobre datos, método, fiabilidad y errores. Si el tribunal utiliza una herramienta para introducir de oficio una fuente o una cuestión decisiva, debe respetarse la contradicción.

También puede producirse una desigualdad inversa: grandes despachos y litigantes institucionales utilizan sistemas avanzados de análisis, mientras personas vulnerables dependen de herramientas generalistas con mayor riesgo de error. La brecha no es solo de acceso, sino de capacidad para detectar alucinaciones y formular *prompts* eficaces. La respuesta institucional debe reforzar la asistencia jurídica, la alfabetización digital y el acceso a fuentes fiables, no trasladar a la ciudadanía la carga de competir tecnológicamente con actores mejor dotados.

La transparencia algorítmica es, en suma, una condición de la contradicción. No se trata de satisfacer curiosidad técnica, sino de impedir que una razón secreta produzca efectos jurídicos.

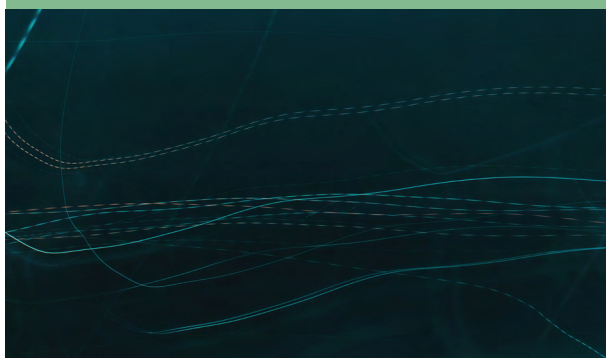
5. TERCER RIESGO: DELEGACIÓN INDEBIDA Y AUTOMATIZACIÓN COMPLACIENTE

5.1. La delegación es un continuo, no un interruptor

La delegación indebida rara vez se presenta como una orden explícita a la máquina para decidir el fallo. Se produce por acumulación de pequeñas cesiones: la herramienta selecciona los documentos relevantes, resume los hechos, identifica las cuestiones, propone la norma, recupera precedentes, ordena los argumentos y redacta la conclusión. Al juez puede quedarle únicamente revisar un texto coherente. Formalmente hay control humano; materialmente el espacio para una deliberación independiente se ha estrechado.

Conviene imaginar una escala. En el nivel más bajo, la IA corrige ortografía o convierte audio en texto. Después clasifica y busca. Más adelante resume, compara y propone estructuras. En los niveles de mayor riesgo formula hipótesis jurídicas, valora consistencia probatoria, recomienda una solución o genera un borrador completo. La frontera no depende de una categoría abstracta, sino del caso, de la intensidad de la influencia y de la posibilidad real de revisión.

“La psicología de la automatización muestra que las personas tienden a aceptar recomendaciones de sistemas percibidos como expertos, especialmente bajo presión temporal”



Una función de resumen puede ser de bajo riesgo cuando sirve para orientarse en un documento ya leído; es de alto riesgo si sustituye su lectura. Un borrador puede ser un apoyo legítimo si el juez ha formado previamente su criterio y reconstruye cada razonamiento; se convierte en delegación cuando la solución emerge de la propuesta y la revisión se limita a corregirla. La misma herramienta cambia de naturaleza según el flujo de trabajo.

5.2. Automatización complaciente, anclaje y fatiga

El Reglamento europeo exige supervisión humana efectiva. La palabra decisiva es «efectiva». La psicología de la automatización muestra que las personas tienden a aceptar recomendaciones de sistemas percibidos como expertos, especialmente

bajo presión temporal. El primer resultado funciona como ancla: obliga a buscar razones para apartarse de él en vez de construir la decisión desde cero.

También existe el llamado sesgo de omisión: el usuario deja de detectar un problema porque el sistema no lo señaló. La herramienta no solo influye mediante lo que dice, sino mediante lo que calla. En documentos extensos, una clasificación negativa puede impedir que un elemento llegue siquiera al campo de atención del juez.

El riesgo de automatización complaciente es particularmente grave en asuntos rutinarios. La repetición genera confianza y reduce la vigilancia; precisamente allí puede ocultarse el caso excepcional. El proceso jurisdiccional existe, entre otras razones, para reconocer la singularidad que la categoría general no capta.

5.3. La escritura judicial no es una fase puramente formal

La utilización de IA únicamente para mejorar la redacción, una vez formado el criterio judicial, plantea una cuestión incómoda: ¿puede sostenerse que la decisión sigue siendo enteramente humana? En muchos casos, escribir no es transcribir una decisión cerrada. La escritura obliga a ordenar hechos, comprobar inferencias, identificar contradicciones y precisar el alcance del fallo. Durante la motivación se descubre a veces que la intuición inicial no se sostiene.

Por eso debe evitarse una separación rígida entre decidir y redactar. Una herramienta que amplía argumentos, selecciona citas o reformula la relación entre hechos y norma puede participar en la construcción de la motivación. Eso no convierte todo apoyo redaccional en ilícito. Exige reconocer que la revisión debe ser más intensa cuanto más sustantiva sea la contribución.

La Instrucción 2/2026 permite, con sistemas autorizados, la generación de borradores a voluntad del juez, libremente modificables, pero prohíbe incorporar contenidos sin validación crítica, completa y personal²⁴. La fórmula es acertada. Validar no es comprobar que el texto suena bien. Es verificar hechos, fuentes, vigencia normativa, pertinencia de precedentes y coherencia con la convicción propia. Supone estar en condiciones de explicar cada párrafo sin apoyo de la herramienta.

5.4. La firma no retiene por sí sola la responsabilidad

Se repite con frecuencia que la responsabilidad seguirá siendo del juez. Es cierto como regla de imputación, pero insuficiente como garantía. Atribuir responsabilidad a una persona que no comprende el proceso tecnológico no recupera el control perdido. La responsabilidad debe ir acompañada de capacidad: acceso a información, formación, tiempo, posibilidad de apartarse y autoridad para suspender el sistema.

Una supervisión humana meramente simbólica puede incluso agravar el problema, porque ofrece una apariencia de legitimidad. La persona aparece como responsable de una decisión configurada por un sistema que no controla. El Reglamento europeo exige que quienes supervisen tengan competencia, formación, autoridad e información suficientes. En el ámbito judicial debe añadirse independencia frente a incentivos organizativos y frente al proveedor.

La responsabilidad no puede descargarse exclusivamente sobre el usuario. Quien adquiere, despliega, diseña y actualiza participa en la cadena causal y debe asumir deberes de documentación, seguridad, auditoría y reparación. El juez responde de la resolución, pero no en solitario de defectos ocultos de un sistema institucionalmente validado.

La literatura reciente sobre responsabilidad por IA confirma una tendencia antropocéntrica: la tecnología no comparece como sujeto al que puedan trasladarse deberes, sino como instrumento dentro de una cadena de diseño, implantación, control y uso. Sotomayor Padilla y otros sintetizan esa evolución en torno a criterios clásicos —diligencia, previsibilidad, causalidad y supervisión— que no desaparecen, aunque la opacidad obligue a adaptar reglas probatorias y deberes de documentación.²⁵ En el ámbito jurisdiccional, la responsabilidad indelegable del juez no excluye la de la institución, la Administración prestacional o el proveedor por fallos de diseño y despliegue. La responsabilidad debe distribuirse a lo largo de la cadena sin diluirse en ella.

5.5. Órganos colegiados y deliberación

En un órgano colegiado, la deliberación previa puede reducir el riesgo de que una IA forme por completo la decisión del ponente. Pero no lo elimina. El borrador estructura el

24. Acuerdo de 28 de enero de 2026, del Pleno del Consejo General del Poder Judicial, por el que se aprueba la Instrucción 2/2026, sobre la utilización de sistemas de inteligencia artificial en el ejercicio de la actividad jurisdiccional, ordinal séptimo, apartados 1, 2 y 4, y ordinal octavo, letra c), BOE núm. 27, de 30 de enero de 2026, BOE-A-2026-2205.

25. P. C. Sotomayor Padilla et al., «Jurisprudencia sobre inteligencia artificial y responsabilidad jurídica», *Ethos Scientific Journal*, vol. 4, n.º 1, 2026, pp. 147-165, especialmente pp. 159-161.

debate posterior, selecciona qué objeciones quedan visibles y puede influir en la adhesión de quienes confían en el trabajo del ponente. Si la contribución de la IA es sustantiva, el resto del tribunal debe conocerla para ejercer un control real.

La deliberación judicial está protegida por su confidencialidad, pero no debe confundirse con opacidad tecnológica. Puede preservarse el secreto de las posiciones individuales y, al mismo tiempo, establecer reglas internas sobre uso de herramientas, trazabilidad y comunicación a los integrantes del órgano. La independencia interna exige que ningún magistrado sea llevado a firmar un texto cuya génesis material desconoce cuándo esa génesis pueda afectar a su valoración.

La clave es funcional: existe delegación indebida cuando la herramienta sustituye una operación que el juez debe realizar personalmente o cuando su influencia no puede ser reconstruida y sometida a crítica. El control humano no se mide por el número de clics, sino por la autonomía cognitiva conservada.

6. GARANTÍAS JURISDICCIONALES FRENTE A LOS RIESGOS DE LA IA

6.1. Derecho al juez y reserva de jurisdicción

El primer límite deriva de la propia atribución constitucional de la potestad jurisdiccional. La ciudadanía tiene derecho a que su controversia sea decidida por el órgano predeterminado por la ley, independiente e imparcial. La IA no puede convertirse en un decisor oculto ni en una fuente normativa informal.

La reserva de jurisdicción no impide utilizar herramientas auxiliares. Impide que la valoración de la prueba, la selección decisoria del Derecho o el juicio de proporcionalidad sean automatizados o delegados. También impide que el resultado de un sistema condicione de hecho la libertad de criterio. La Instrucción 2/2026 recoge expresamente estas prohibiciones²⁶.

La garantía exige que el juez pueda apartarse de la recomendación sin penalización, sin necesidad de justificar internamente por qué «desobedece» a la máquina y sin que el sistema convierta la discrepancia en una anomalía de desempeño. Una herramienta que registra sistemáticamente quién se aparta para evaluar productividad puede generar presión incompatible con la independencia.

6.2. Independencia judicial y dependencia tecnológica

La independencia no se protege solo frente a órdenes políticas o jerárquicas. También puede verse comprometida por infraestructuras que delimitan silenciosamente las opcio-

26. Acuerdo de 28 de enero de 2026, del Pleno del Consejo General del Poder Judicial, por el que se aprueba la Instrucción 2/2026, sobre la utilización de sistemas de inteligencia artificial en el ejercicio de la actividad jurisdiccional, ordinal cuarto, letras a), b) y d), y ordinal octavo, letras a) y b), BOE núm. 27, de 30 de enero de 2026, BOE-A-2026-2205.

nes. Si un proveedor decide qué jurisprudencia se indexa, cómo se ordena, cuándo se actualiza el modelo o qué restricciones introduce, participa indirectamente en el entorno de decisión judicial.

De ahí la importancia de la soberanía de datos, la interoperabilidad, la posibilidad de auditoría y la reversibilidad contractual. La CEPEJ recomienda avanzar hacia soluciones especializadas y, cuando sea adecuado, despliegues locales o bajo control soberano²⁷. No siempre será necesario un modelo íntegramente público, pero sí contratos que aseguren acceso técnico, conservación de registros, notificación de cambios, prohibición de reutilizar datos, continuidad del servicio y capacidad de migración.

Sin conocimiento técnico propio, el Poder Judicial dependerá de la explicación del proveedor. La auditoría no puede ser exclusivamente contractual ni realizada por quien vende el producto: requiere equipos públicos interdisciplinarios y evaluación independiente.

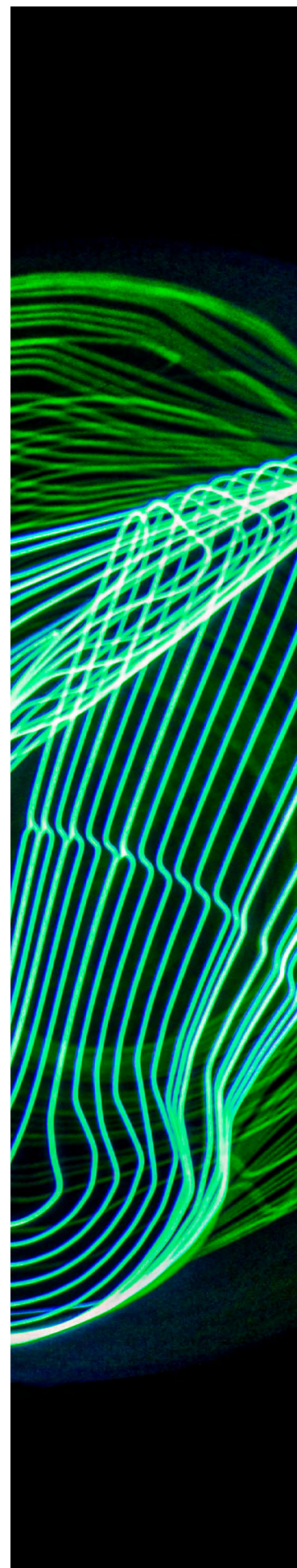
6.3. Igualdad de armas y contradicción

Cuando un resultado algorítmico influya materialmente en el proceso, las partes deben conocer su existencia y disponer de una oportunidad razonable para cuestionarlo. La información necesaria dependerá de la función del sistema, pero debe permitir discutir los datos, la finalidad, la tasa de error, la población de validación, los factores relevantes y el modo en que el resultado fue utilizado.

En el proceso penal, la defensa no puede quedar obligada a refutar una puntuación cuyo funcionamiento se protege por secreto comercial. En el contencioso-administrativo, la Administración no puede limitarse a aportar el resultado de una aplicación sin explicar cómo tradujo la norma. En lo social o civil, una pericial construida mediante IA debe someterse a las mismas exigencias de fiabilidad, metodología y contradicción que cualquier otra prueba técnica.

La contradicción puede requerir peritos, acceso controlado a documentación o pruebas sobre casos simulados. Los costes no deberían recaer desproporcionadamente

27. Comisión Europea para la Eficacia de la Justicia (CEPEJ), *Guidelines on the Use of Generative Artificial Intelligence for Courts*, documento CEPEJ(2025)18Final, Consejo de Europa, Estrasburgo, 19 de diciembre de 2025



sobre la parte vulnerable. Cuando el sistema ha sido elegido por un poder público, corresponde a este asegurar las condiciones de impugnación.

6.4. Motivación, comprensibilidad y razones reales

El deber de motivación cumple varias funciones: permite conocer por qué se decide, disciplina el razonamiento, evita arbitrariedad y hace posible el recurso. La IA puede mejorar claridad y estructura, pero también producir una motivación estandarizada, abundante y vacía.

Una garantía mínima es que la resolución no se limite a reproducir puntuaciones o recomendaciones. Debe explicar cómo se relacionan con los hechos probados y por qué resultan jurídicamente relevantes. Si el sistema aportó una probabilidad, ha de aclararse que no constituye certeza individual y qué peso se le concedió. Si la IA localizó precedentes, debe comprobarse su autenticidad y explicar su aplicabilidad.

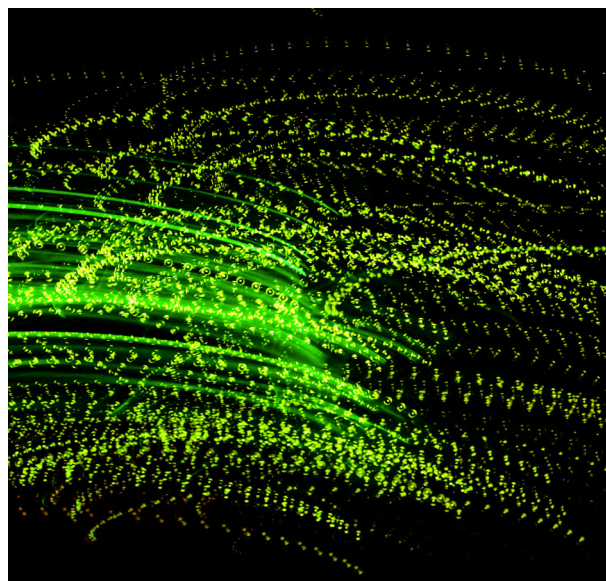
La motivación debe expresar razones humanas asumidas. No es necesario relatar el proceso psicológico completo ni revelar deliberaciones protegidas, pero sí evitar que la decisión se apoye en factores inaccesibles. La frase «se ha tenido en cuenta el resultado del sistema» es insuficiente si ese resultado fue relevante y no se puede comprender.

6.5. Prueba, intermediación y valoración individualizada

La IA puede ayudar a organizar prueba documental o detectar relaciones en grandes volúmenes de información. Pero no debe atribuir credibilidad a personas mediante análisis facial, voz, emoción o conducta no verbal. Estas tecnologías carecen de una base científica suficientemente robusta para inferir estados internos y pueden re-

producir estereotipos. El Reglamento europeo contiene prohibiciones y severas restricciones en materias biométricas y de reconocimiento de emociones.

La intermediación no significa intuición infalible, pero protege la relación directa con la prueba y la posibilidad de contradicción. Un sistema que transforma un testimonio en etiquetas de veracidad añade una capa opaca incompatible con la valoración racional. La tecnología puede transcribir; no debe decidir quién dice la verdad.



“La IA puede ayudar a organizar prueba documental o detectar relaciones en grandes volúmenes de información. Pero no debe atribuir credibilidad a personas mediante análisis facial, voz, emoción o conducta no verbal”

La valoración individualizada es igualmente esencial en medidas cautelares, ejecución penal, protección de menores y capacidad. Las herramientas de riesgo describen patrones poblacionales. No pueden sustituir la consideración de circunstancias concretas ni convertir pertenencia a un grupo en fundamento de restricción.

6.6. Presunción de inocencia y prevención de perfilado

Los sistemas predictivos desplazan la atención del hecho probado hacia la probabilidad de conducta futura. Esa lógica puede ser legítima en ámbitos muy delimitados, pero entraña el riesgo de sancionar a la persona por lo que estadísticamente se espera de quienes se le parecen.

La Instrucción 2/2026 prohíbe el perfilado, la predicción de comportamientos, la evaluación de riesgos o la clasificación de sujetos fuera de supuestos expresamente autorizados²⁸. El Reglamento europeo califica como alto riesgo diversos sistemas en garantía del cumplimiento del Derecho y administración de justicia, y prohíbe determinadas prácticas²⁹.

La garantía jurisdiccional exige base legal clara, finalidad delimitada, necesidad, proporcionalidad, validación local y prohibición de usar la puntuación como único o principal fundamento. Debe evitarse que factores socioeconómicos funcionen como sustitutos de culpabilidad o peligrosidad.

6.7. Información, trazabilidad y derecho a explicación

No todo uso menor de IA requiere comunicación procesal. Una corrección ortográfica o una búsqueda auxiliar sin incidencia material puede equipararse a otras herramientas. Pero cuando el sistema influye en la selección de hechos, valoración de riesgos, admisión de prueba, determinación de medidas o contenido sustantivo de la motivación, la transparencia debe aumentar.

El artículo 86 del Reglamento europeo reconoce, en determinados supuestos, el derecho a obtener explicaciones claras y significativas sobre el papel del sistema de alto riesgo y los principales elementos de la decisión³⁰. La jurisprudencia del TJUE sobre decisiones automatizadas refuerza una explicación comprensible, no reducida a revelar fórmulas³¹.

28. Acuerdo de 28 de enero de 2026, del Pleno del Consejo General del Poder Judicial, por el que se aprueba la Instrucción 2/2026, sobre la utilización de sistemas de inteligencia artificial en el ejercicio de la actividad jurisdiccional, apartados octavo.1, 2 y 4, y octavo.e), BOE núm. 27, de 30 de enero de 2026, BOE-A-2026-2205

29. Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial, arts. 5 y 6.2 y anexo III, apartados 6 —garantía del cumplimiento del Derecho— y 8 —administración de justicia y procesos democráticos—, DOUE L, de 12 de julio de 2024, ELI: <http://data.europa.eu/eli/reg/2024/1689/oj>. El artículo 5 recoge las prácticas prohibidas; el artículo 6.2, en relación con el anexo III, determina los sistemas considerados de alto riesgo.

30. Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial, artículo 86 del Reglamento (UE) 2024/1689 y anexo III, apartados 6 —garantía del cumplimiento del Derecho— y 8 —administración de justicia y procesos democráticos—, DOUE L, de 12 de julio de 2024, ELI: <http://data.europa.eu/eli/reg/2024/1689/oj>.

31. STJUE (Sala Primera) de 27 de febrero de 2025, asunto C-203/22, *Dun & Bradstreet Austria*, ECLI:EU:C:2025:117, apartados 55-66, especialmente apartado 59. El TJUE exige una explicación que permita a la persona afectada comprender y cuestionar la decisión automatizada; no basta con comunicar una fórmula matemática compleja o el algoritmo, pues ello no constituye necesariamente una explicación concisa e inteligible.

Propongo un criterio de materialidad: debe informarse cuando una persona razonable pudiera considerar que la intervención de la IA tuvo capacidad real para afectar al resultado o a las posibilidades de defensa. La información puede incorporarse al expediente mediante una nota normalizada: sistema, versión, finalidad, datos tratados, resultado utilizado y verificación humana.

La trazabilidad debe ser proporcionada y respetar el secreto deliberativo. No se trata de conservar cada borrador mental ni cada consulta irrelevante, sino de permitir reconstruir la intervención tecnológica que haya producido efectos jurídicos.

6.8. Derecho al recurso y revisión humana

La revisión exige que el órgano superior pueda examinar tanto la decisión como la influencia material de la herramienta. Si los registros se destruyen o el sistema cambia de versión, la revisión queda vacía. Los contratos y protocolos deben asegurar conservación durante plazos compatibles con recursos y responsabilidades.

La persona afectada debe poder solicitar intervención humana cuando la decisión haya sido automatizada o fuertemente condicionada. Esa intervención no puede consistir en que otro usuario confirme el mismo resultado mirando la misma pantalla. Debe incluir autoridad para revisar datos, corregir errores y apartarse del sistema.

Cuando no sea posible ofrecer una explicación suficiente o reproducir el resultado, debe cuestionarse la admisibilidad de su uso en una decisión desfavorable. La carga de la opacidad no puede recaer sobre quien sufre sus efectos.

6.9. Protección de datos y confidencialidad jurisdiccional

Los expedientes contienen datos de especial sensibilidad: salud, violencia, menores, antecedentes, situación económica, relaciones familiares y comunicaciones privadas. Introducirlos en una herramienta no autorizada puede implicar comunicación a un tercero, transferencia, conservación o reutilización incompatible.

La anonimización no siempre resuelve el problema. La combinación de hechos puede permitir reidentificación; además, el contenido puede estar sometido a secreto, aunque no incluya nombres. Deben aplicarse minimización, limitación de finalidad, seguridad, control de acceso, cifrado, registro y prohibición de entrenamiento con datos jurisdiccionales salvo base jurídica y garantías específicas.

El riesgo inmediato no es solo una resolución errónea, sino también la pérdida de control sobre datos del procedimiento. Que una aplicación sea de pago, esté instalada en un equipo corporativo u ofrezca una opción para excluir entrenamiento no equivale a autorización jurisdiccional. La Instrucción 2/2026 permite únicamente sistemas facilitados por las Administraciones competentes o por el CGPJ, y exige auditoría del Consejo para los ofrecidos por aquellas³².

32. Acuerdo de 28 de enero de 2026, del Pleno del Consejo General del Poder Judicial, por el que se aprueba la Instrucción 2/2026, sobre la utilización de sistemas de inteligencia artificial en el ejercicio de la actividad jurisdiccional, apartados ordinal quinto, apartados 1 y 2, y ordinal noveno, apartado 1), BOE núm. 27, de 30 de enero de 2026, BOE-A-2026-2205.

7. EL MARCO NORMATIVO EUROPEO Y ESPAÑOL

7.1. El Reglamento (UE) 2024/1689

El marco europeo no debe leerse como una sucesión inconexa de reglamentos. Fernando H. Llano Alonso propone comprenderlo como un nuevo *corpus iuris digitalis*: un cuerpo normativo en formación que intenta responder a la revolución tecnológica sin renunciar al humanismo jurídico.³³ En él convergen la Carta de Derechos Fundamentales, el RGPD, el Reglamento de IA, las normas sobre servicios y mercados digitales, la gobernanza de datos, la ciberseguridad y el Convenio Marco del Consejo de Europa sobre IA. Su unidad no es simplemente temática. La proporciona una misma pretensión: someter el poder digital a dignidad, libertad, igualdad, responsabilidad y tutela efectiva.

La noción resulta útil para evitar dos errores. El primero consiste en creer que el cumplimiento del Reglamento de IA agota la legalidad del sistema. Una herramienta puede superar la evaluación de conformidad y vulnerar, por su uso concreto, protección de datos, defensa, igualdad o independencia judicial. El segundo es imaginar que la ética comienza donde termina el Derecho. En el *corpus iuris digitalis* europeo, principios humanistas y obligaciones técnicas se comunican: la dignidad se traduce en prohibiciones y control humano; la igualdad, en gobernanza de datos y auditoría de sesgos; la autonomía, en información, explicación e impugnación; la responsabilidad, en registros, supervisión y remedios.

Llano advierte, al mismo tiempo, del riesgo de deshumanización del Derecho en la era digital. No se produce únicamente cuando una máquina dicta la resolución. Comienza antes: cuando la eficiencia desplaza a la justicia como criterio de diseño, cuando la persona se convierte en dato y cuando el jurista renuncia al juicio crítico porque la respuesta técnica parece inevitable. La respuesta no puede ser nostalgia analógica, sino un humanismo jurídico capaz de gobernar la técnica.

El Reglamento europeo adopta un enfoque basado en el riesgo³⁴. El anexo III incluye como sistemas de alto riesgo los destinados a ayudar a una autoridad judicial en la investigación e interpretación de hechos y Derecho y en la aplicación del Derecho a un conjunto concreto de hechos. No toda herramienta utilizada en un tribunal queda automáticamente incluida: el artículo 6.3 contempla excepciones para tareas procedimentales limitadas, preparatorias o que no influyan sustancialmente en el resultado, aunque el perfilado mantiene la consideración de alto riesgo.

La clasificación depende de la finalidad prevista y del uso real. Un gestor de agenda no es equiparable a un sistema que propone la solución de un recurso. Una herramienta inicialmente administrativa puede convertirse en alto riesgo si se modifica su finalidad o se integra en el razonamiento decisorio.

33. F. H. Llano Alonso, *El jurista ante los retos de la revolución tecnológica 4.0*, en A. M. Marcos del Cano (coord.), *Inteligencia artificial. Desafíos éticos, sociales y jurídicos*, Dykinson, Madrid, 2026, pp. 21-36, especialmente sus apartados sobre la deshumanización del Derecho y el nuevo *corpus iuris digitalis*.

34. Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial, arts. 5 y 6.2 y anexo III, apartados 6 —garantía del cumplimiento del Derecho— y 8 —administración de justicia y procesos democráticos—, DOUE L, de 12 de julio de 2024, ELI: <http://data.europa.eu/eli/reg/2024/1689/oj>.

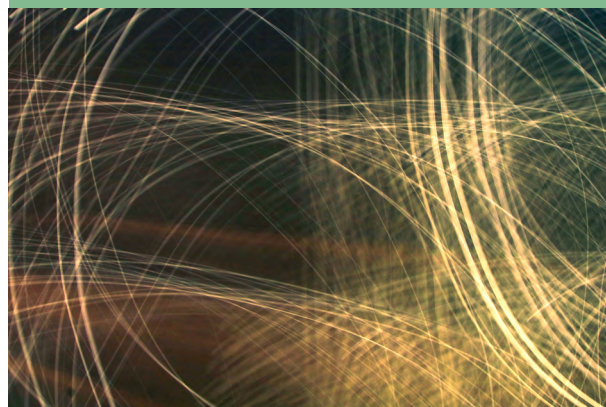
Para los sistemas de alto riesgo, el Reglamento exige gestión continua de riesgos; datos pertinentes, representativos y sometidos a gobernanza; documentación y registros; transparencia suficiente para interpretar los resultados; precisión, solidez y ciberseguridad; supervisión humana efectiva; obligaciones del responsable del despliegue; y, para determinados organismos públicos, evaluación de impacto sobre derechos fundamentales. El artículo 27 obliga a identificar personas y colectivos afectados, riesgos específicos, medidas de supervisión, gobernanza interna y mecanismos de reclamación.

La fecha general de aplicación del Reglamento es el 2 de agosto de 2026. No obstante, tras la modificación introducida por el Reglamento (UE) 2026/1744, de 8 de julio de 2026, las secciones 1, 2 y 3 del capítulo III, con excepción del artículo 6.5, serán aplicables desde el 2 de diciembre de 2027 respecto de los sistemas de IA clasificados como de alto riesgo en virtud del artículo 6.2 y del anexo III, y desde el 2 de agosto de 2028 respecto de los sistemas clasificados como de alto riesgo en virtud del artículo 6.1 y del anexo I. Este calendario se entiende sin perjuicio de la aplicación de las garantías constitucionales y procesales y de las restantes normas que resulten aplicables.³⁵

7.2. La Instrucción 2/2026 del CGPJ³⁶

La Instrucción constituye el marco operativo inmediato para la carrera judicial española. Sus ejes son control humano real, no sustitución, responsabilidad judicial, independencia, derechos fundamentales,

“La clave no es solo si el documento está formalmente anonimizado, sino si procede del expediente y está sometido a deberes de confidencialidad”



confidencialidad, prevención de sesgos, proporcionalidad y formación.

Solo permite utilizar en actividad jurisdiccional sistemas facilitados por las Administraciones competentes o por el CGPJ; los primeros deben ser objeto de control de calidad y auditoría por el Consejo. Prohíbe sustitución, automatización o delegación de la decisión, valoración de hechos y pruebas e interpretación y aplicación del Derecho; prohíbe resultados que condicionen independencia; y excluye contenidos generados sin validación crítica, completa y personal.

La excepción para preparación o estudio con fuentes abiertas permite utilizar herra-

35. Reglamento (UE) 2026/1744 del Parlamento Europeo y del Consejo, de 8 de julio de 2026, por el que se modifican los Reglamentos (UE) 2024/1689, (UE) 2018/1139 y (UE) 2023/1230 en lo que respecta a la simplificación de la aplicación de normas armonizadas en materia de inteligencia artificial (Ómnibus digital sobre IA), DO L, 2026/1744, de 24 de julio de 2026.

36. Acuerdo de 28 de enero de 2026, del Pleno del Consejo General del Poder Judicial, por el que se aprueba la Instrucción 2/2026, sobre la utilización de sistemas de inteligencia artificial en el ejercicio de la actividad jurisdiccional, apartados ordinales cuarto, quinto, séptimo, octavo y noveno», BOE núm. 27, de 30 de enero de 2026, BOE-A-2026-2205.

mientas no oficialmente habilitadas para resumir o traducir materiales públicos, pero no para introducir información obtenida en el ejercicio jurisdiccional. La clave no es solo si el documento está formalmente anonimizado, sino si procede del expediente y está sometido a deberes de confidencialidad.

La Instrucción debe interpretarse como un suelo de garantías, no como una autorización automática de cualquier uso posible en un sistema oficial. Que una herramienta haya sido facilitada elimina un primer obstáculo; no dispensa de valorar finalidad, proporcionalidad, fiabilidad y efecto sobre el proceso concreto.

7.3. Habilitación legal, reserva de humanidad y supervisión institucional

El Reglamento de IA establece requisitos de conformidad y uso, pero no es por sí solo la norma que atribuye potestad para introducir cualquier sistema en el ejercicio jurisdiccional. Esta distinción, desarrollada por Lorenzo Cotino Hueso, es esencial: la conformidad europea de un producto no sustituye la habilitación legal que puede exigir una injerencia en derechos ni define qué operaciones pertenecen al núcleo reservado a la persona que juzga.³⁷

Cotino recupera la expresión reserva de humanidad, acuñada en la doctrina española por Juli Ponce, para delimitar espacios en los que deben preservarse el criterio, la sensibilidad y la responsabilidad humanos. En la jurisdicción, esa reserva se proyecta sobre la verificación de los hechos, la valoración de la prueba, la interpretación del Derecho y el pronunciamiento. La Instrucción 2/2026 ha positivizado operativamente buena parte de esta idea al prohibir

la sustitución, automatización o delegación de esas funciones.

La reserva de humanidad no significa reserva de ineficiencia ni prohibición de asistencia. Impone una diferencia cualitativa entre preparar y decidir. Una herramienta puede facilitar materiales, detectar relaciones o proponer un borrador; pero cuanto más se acerque al núcleo decisorio, más densa debe ser la base legal, más intensa la contradicción y menos admisible la opacidad.

El mismo trabajo propone garantías concretas: delimitación del uso y de los datos, usuarios autorizados, duración, seudonimización, conservación y destrucción, auditorías externas, transparencia de la lógica y trazabilidad de la correspondencia entre salida automatizada y decisión final. Nuestro protocolo coincide sustancialmente con esa propuesta y añade la perspectiva de la interacción cognitiva y la igualdad procesal.

Por último, la supervisión no puede ignorar la separación de poderes. Cotino sostiene que, para los sistemas jurisdiccionales de alto riesgo, el CGPJ debe ocupar la posición de autoridad de vigilancia por las exigencias de independencia institucional, sin perjuicio de la cooperación con la AEPD, la AESIA y el CTEAJE en sus respectivos ámbitos. La tesis merece atención porque la independencia frente al Ejecutivo alcanza también a quien inspecciona el sistema, accede a sus registros, exige cambios o decide su suspensión. A la vez, la legitimidad de esa solución depende de que el Consejo cuente con capacidad técnica, recursos, procedimientos transparentes y participación suficiente: la competencia formal sin medios no produce supervisión efectiva.

37. L. Cotino Hueso, «El uso jurisdiccional de la inteligencia artificial: habilitación legal, garantías necesarias y la supervisión por el CGPJ», *Actualidad Jurídica Iberoamericana*, n.º 21, 2024, pp. 494-527, especialmente pp. 497-498 y 515-527.

7.4. CTEAJE y CEPEJ

La Política de uso de la IA en la Administración de Justicia aprobada por el CTEAJE en 2024 exige sistemas accesibles, comprensibles y auditables; prevención de sesgos; evaluaciones periódicas; protección de datos; información a la ciudadanía; revisión humana y derecho a objetar. Su importancia reside en integrar la IA dentro de la arquitectura de derechos y seguridad de la Administración de Justicia.

La Carta Ética de la CEPEJ de 2018 articuló cinco principios: derechos fundamentales, no discriminación, calidad y seguridad, transparencia e imparcialidad, y control por el usuario. Las reflexiones y directrices de 2025 sobre IA generativa desarrollan esos principios: acceso permanente a un juez humano, carácter no vinculante de los resultados, herramientas especializadas y seguras, trazabilidad, evaluación ex ante, formación, verificación de fuentes y responsabilidad pública.

7.5. Jurisprudencia y control judicial del algoritmo

BOSCO demuestra que la automatización administrativa puede ser sometida a transparencia judicial reforzada. *Dun & Bradstreet Austria* enseña que explicar no equivale a entregar un algoritmo, sino a hacer comprensible el procedimiento y los factores relevantes. *SCHUFA* había advertido ya que una puntuación puede constituir, en la práctica, una decisión automatizada cuando el tercero le atribuye un papel determinante.

Estos precedentes permiten anticipar una doctrina general: cuanto mayor sea el impacto del sistema sobre derechos, mayor debe ser la intensidad de la explicación, la auditabilidad y la revisión. La opacidad contractual o técnica no puede vaciar el derecho a la tutela judicial efectiva.

7.5.1. *SCHUFA*: una recomendación puede ser materialmente una decisión

En la sentencia de 7 de diciembre de 2023, *SCHUFA Holding (Scoring)*, el TJUE interpretó de forma amplia el artículo 22.1 RGPD³⁸. La generación automatizada de una puntua-

38. STJUE (Sala Primera) de 7 de diciembre de 2023, asunto C-634/21, *SCHUFA Holding (Scoring)*, ECLI:EU:C:2023:957, especialmente apartados 60-73 y fallo. El Tribunal consideró que la elaboración automatizada de una puntuación de solvencia puede constituir una «decisión individual automatizada» del artículo 22.1 RGPD cuando un tercero atribuye a esa puntuación un papel determinante en la decisión contractual.



ción de solvencia podía constituir por sí misma una «decisión» cuando las entidades que recibían el resultado le atribuían un papel determinante al conceder o denegar crédito.³⁹ Lo relevante no era la distribución formal de competencias entre la empresa que calculaba el score y el banco que firmaba la negativa, sino la influencia real del resultado automatizado.

Esta mirada material es trasladable, con las debidas cautelas, al uso judicial. No basta afirmar que la herramienta «solo recomienda» si su salida estructura de hecho la respuesta. La cuestión debe examinarse en el contexto de uso: frecuencia de adhesión, forma de presentación, posibilidad de apartamiento, tiempo disponible, información alternativa y consecuencias de la discrepancia. Un sistema consultivo puede ser decisorio en la práctica; una firma humana no rompe necesariamente la cadena automatizada.

7.5.2. *Dun & Bradstreet Austria*: explicar no es exhibir una fórmula

La sentencia de 27 de febrero de 2025, C-203/22, dio un paso adicional. El derecho a «información significativa sobre la lógica aplicada» exige una explicación que permita a la persona comprender y cuestionar la decisión. No obliga a entregar fórmulas complejas o el algoritmo completo cuando esa entrega resulte ininteligible; requiere describir, de forma concisa, transparente y accesible, el procedimiento y los principios concretamente aplicados. El Tribunal consideró que podían utilizarse explicaciones contrafactuales, mostrando qué variación de los datos habría producido un resultado diferente.⁴⁰

El pronunciamiento evita dos reduccionismos. El primero identifica transparencia con apertura del código. El segundo identifica secreto empresarial con silencio. Si la información afecta a derechos de terceros o a secretos comerciales, debe ponerse a disposición de la autoridad o del tribunal competente para que realice la ponderación; el responsable no puede resolver unilateralmente el conflicto ocultando la lógica. Para la Justicia, la lección es clara: una explicación útil se diseña desde la posición de quien debe defenderse, no desde la comodidad de quien produce el sistema.

7.5.3. *BOSCO*: del derecho de acceso al control democrático del código

En *BOSCO*, STS de 11 de septiembre de 2025, la Sala Tercera reconoció a la Fundación Civio el derecho de acceder al código fuente de la aplicación que comprobaba los requisitos del bono social eléctrico.⁴¹ La resolución es especialmente importante por el objeto —una prestación dirigida a personas vulnerables— y porque rechaza que propiedad intelectual y seguridad actúen como límites automáticos. La Administración debía justificar el daño de manera concreta y ponderarlo con el intenso interés público en comprobar si el programa traducía correctamente las condiciones legales.

El acceso al código cumple aquí una función democrática y probatoria: permite verificar si la norma jurídica se convirtió fielmente en regla computacional, identificar errores y fiscalizar una infraestructura pública. Pero el propio avance de la técnica obliga a prolongar la doctrina. En un sistema determinista, el código puede revelar la operación. En aprendizaje automático deben añadirse datos, variables, umbrales y validaciones; en

39. STJUE (Sala Primera) de 7 de diciembre de 2023, asunto C-634/21, SCHUFA Holding (Scoring), ECLI:EU:C:2023:957, aps. 43-73.

40. TJUE, *Dun & Bradstreet Austria*, C-203/22, en particular aps. 59-68 y fallo.

41. STS, Sala Tercera, 11 de septiembre de 2025, rec. 7878/2024, ECLI:ES:TS:2025:3826.

IA generativa, versión del modelo, instrucciones, fuentes recuperadas y parámetros. El principio que sobrevive a las arquitecturas es el de **auditabilidad suficiente para el control que se pretende ejercer**.

7.5.4. Estrasburgo: garantías procedimentales para el poder tecnológico

El TEDH todavía no ha construido una doctrina general sobre IA utilizada por los jueces. Conviene decirlo expresamente. Sí ofrece, sin embargo, principios aplicables al ejercicio de poder público mediado por tecnologías complejas. En *Big Brother Watch y otros c. Reino Unido*, la Gran Sala exigió «*salvaguardas de extremo a extremo*» en los regímenes de interceptación masiva: evaluación de necesidad y proporcionalidad en cada fase, autorización independiente, supervisión y control posterior.⁴² La analogía no reside en equiparar vigilancia e IA judicial, sino en una idea de garantía: cuanto más invisible y escalable es el poder técnico, menos puede descansar su legitimidad en una autorización inicial genérica.

En *Glukhin c. Rusia*, el Tribunal calificó el reconocimiento facial como tecnología altamente intrusiva y apreció vulneración de los artículos 8 y 10 CEDH por su uso para identificar y localizar a un manifestante pacífico sin garantías suficientes ni necesidad acreditada.⁴³ El caso recuerda que la eficiencia identificativa no sustituye el examen de legalidad, necesidad y proporcionalidad, y que la dificultad de la persona para demostrar el uso de la tecnología debe ser tenida en cuenta cuando el Estado controla los registros.

Para el proceso equitativo resultan igualmente relevantes las líneas de *Baka c. Hungría* y *Guðmundur Andri Ástráðsson c. Islandia*. La primera protege la independencia judicial frente a medidas con efecto disuasorio; la segunda conecta el «tribunal establecido por la ley» con un procedimiento de designación capaz de preservar la confianza pública.⁴⁴ Ninguna se refiere a algoritmos. Ambas confirman, no obstante, que la independencia se valora mediante condiciones institucionales efectivas y no solo por la ausencia formal de órdenes en un litigio concreto. Esa misma aproximación funcional debe inspirar el examen de infraestructuras capaces de condicionar de forma sistemática cómo conocen y deciden los tribunales.

7.5.5. Hacia un estándar jurisprudencial integrado

De este conjunto heterogéneo de resoluciones puede extraerse —como propuesta doctrinal, no como regla ya formulada por un único tribunal— un estándar de siete exigencias:

1. atender a la influencia material del sistema y no a su etiqueta de «ayuda» o «recomendación»;
2. asegurar base legal, finalidad legítima, necesidad y proporcionalidad;
3. informar de la intervención de la IA cuando sea relevante para derechos;

42. TEDH (Gran Sala), *Big Brother Watch y otros c. Reino Unido*, 25 de mayo de 2021, especialmente aps. 350 y ss.

43. TEDH, *Glukhin c. Rusia*, 4 de julio de 2023, aps. 73-90.

44. TEDH (Gran Sala), *Baka c. Hungría*, 23 de junio de 2016; y *Guðmundur Andri Ástráðsson c. Islandia*, 1 de diciembre de 2020.

4. facilitar una explicación funcional que permita comprender e impugnar;
5. preservar registros y acceso suficiente para auditoría y prueba;
6. garantizar revisión humana con competencia, autoridad y tiempo reales; y
7. ofrecer un remedio efectivo, incluida la posibilidad de excluir el resultado cuando su opacidad haga imposible la contradicción.

Estas garantías no convierten cualquier IA en admisible. Permiten formular la pregunta previa: si el sistema no puede satisfacerlas por su diseño, por el secreto que lo rodea o por el contexto organizativo, ¿es legítimo que intervenga en una decisión jurisdiccional? En determinados usos, la respuesta deberá ser negativa.

8. PROPUESTA DE PROTOCOLO DE GARANTÍAS PARA EL USO JUDICIAL DE IA

La regulación abstracta necesita traducirse en un procedimiento institucional. Propongo doce garantías acumulativas.

1. Finalidad legítima, necesidad y proporcionalidad

Antes de adquirir o desplegar una herramienta debe definirse el problema concreto. No basta afirmar que «mejorará la eficiencia». Debe compararse con alternativas menos intrusivas y explicar por qué la IA aporta valor. Un modelo complejo no es preferible a una regla transparente si ambos ofrecen rendimiento similar.

2. Clasificación previa del riesgo y delimitación del uso

Cada sistema debe contar con una ficha que identifique finalidad, usuarios, población afectada, tipo de datos, decisiones en las que interviene y nivel de influencia permitido. Deben prohibirse usos secundarios no evaluados. Cambiar de contexto exige nueva valoración.

3. Autorización institucional y entorno seguro

Solo deben utilizarse sistemas oficialmente facilitados y auditados cuando se trate información jurisdiccional. El entorno debe impedir reutilización para entrenamiento no autorizada, accesos de terceros, transferencias incompatibles y conservación indefinida. La configuración por defecto debe ser la más protectora.

4. Evaluación de impacto sobre derechos fundamentales y protección de datos

La evaluación debe realizarse antes del despliegue, con participación de especialistas jurídicos y técnicos, representantes de usuarios y, cuando proceda, colectivos afectados. Debe analizar igualdad, defensa, privacidad, libertad, acceso a la justicia y efectos interseccionales. No puede ser un documento estático: debe actualizarse ante cambios relevantes.

5. Calidad y gobernanza de datos

Han de documentarse origen, cobertura, período, criterios de inclusión, etiquetas, variables sustitutivas y limitaciones. Las pruebas deben incluir subgrupos y entorno territorial

y lingüístico de uso. Los datos judiciales no son neutrales: deben interpretarse a la luz de cómo fueron producidos.

6. Auditoría independiente y periódica

La evaluación del proveedor es insuficiente. Se necesitan auditorías técnicas y jurídicas independientes antes del uso y durante su vida útil. Deben examinar precisión, tasas de error por grupos, seguridad, robustez, explicaciones, deriva del modelo y cumplimiento de finalidad. Los resultados esenciales deben hacerse públicos.

7. Transparencia funcional y documentación accesible

El usuario judicial necesita instrucciones claras sobre capacidades, límites, datos adecuados, errores previsibles y forma correcta de interpretar la salida. Las partes necesitan información suficiente para la contradicción cuando el sistema influya materialmente. La transparencia debe adaptarse al destinatario: técnica para peritos, jurídica para tribunales y comprensible para ciudadanía.

8. Supervisión humana por diseño

La interfaz no debe inducir aceptación automática. Conviene separar la formación inicial del criterio humano de la visualización de la recomendación en tareas decisorias, mostrar incertidumbre, ofrecer alternativas y facilitar la consulta de fuentes. El juez debe poder ignorar, corregir o detener el sistema sin consecuencias adversas.

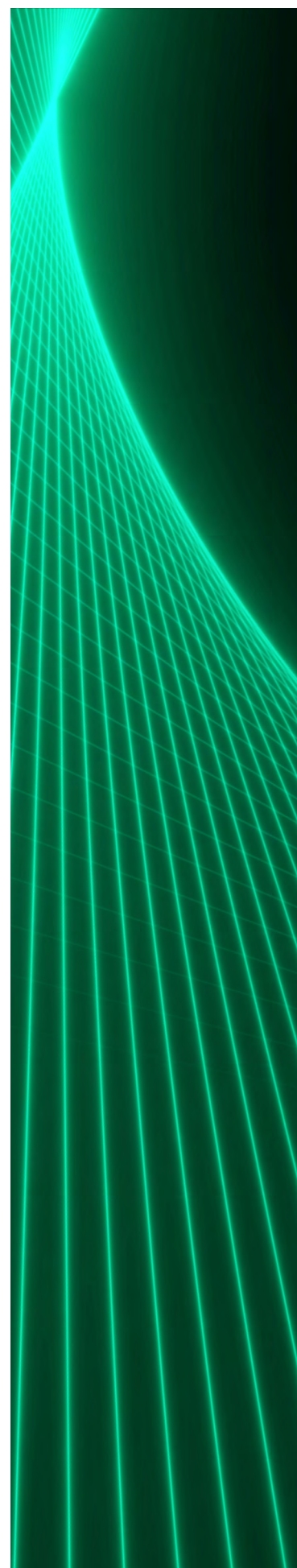
Una práctica recomendable consiste en formular una hipótesis provisional antes de consultar una herramienta sustantiva. Después debe contrastarse la salida con la hipótesis, buscar activamente razones en contra y verificar fuentes. Así se reduce el anclaje.

9. Regla de verificación reforzada

Toda cita, norma, dato y referencia generada debe comprobarse en una fuente oficial o auténtica. Los resúmenes de expedientes deben contrastarse con los documentos originales antes de decidir. Cuando no sea posible verificar una afirmación, no debe incorporarse.

10. Trazabilidad, conservación y comunicación material

Debe registrarse el sistema y versión, fecha, finalidad, fuentes o documentos procesados, salida relevante y controles realizados, con límites que protejan el secreto deliberativo. Los registros deben conservarse durante el tiempo nece-



sario para recursos, auditorías e incidentes. Cuando la influencia sea material, las partes deben ser informadas mediante una fórmula normalizada.

11. Incidentes, reclamación y reversibilidad

Debe existir un canal para comunicar errores, sesgos y usos indebidos, con protección frente a represalias. Los incidentes graves han de provocar revisión y, si procede, suspensión. El sistema debe poder desconectarse sin paralizar la actividad judicial. La dependencia de un proveedor no puede hacer técnicamente imposible volver al procedimiento anterior.

12. Formación, tiempo y cultura institucional

La alfabetización en IA no consiste en aprender a escribir prompts. Incluye comprender límites estadísticos, sesgos, protección de datos, seguridad, alucinaciones, deberes de transparencia y técnicas de verificación. También exige tiempo. Sin una carga de trabajo compatible, el control humano efectivo se convierte en una declaración vacía.

La formación debe incluir fallos reales, no solo demostraciones exitosas, y la cultura institucional debe premiar la detección temprana de errores.

9. GUÍA PRÁCTICA PARA EL JUEZ O JUEZA QUE UTILIZA UNA HERRAMIENTA DE IA

Antes de utilizarla, deberían formularse siete preguntas:

1. **¿Está autorizada para esta tarea y para estos datos?** Que aparezca instalada o accesible no basta.
2. **¿Qué función concreta voy a delegar?** Si afecta a hechos, prueba, interpretación o decisión, el riesgo es máximo.
3. **¿He estudiado suficientemente el asunto para detectar una respuesta incorrecta?** No puede supervisarse lo que no se domina.
4. **¿Conozco las fuentes y limitaciones del sistema?** Una respuesta sin fuente verificable es solo una hipótesis.
5. **¿Podría explicar a las partes cómo influyó la herramienta?** Si no, quizá la influencia sea incompatible con la contradicción.
6. **¿Estoy aceptando el resultado por falta de tiempo?** La urgencia es un indicador de riesgo, no una justificación.
7. **¿Conservaré la autonomía para apartarme?** Si la respuesta negativa genera presión o carga adicional desproporcionada, la supervisión no es efectiva.

Después del uso, el juez debe poder reconstruir la decisión sin depender de la herramienta. El texto final ha de expresar su propio razonamiento, incluir únicamente afirmaciones verificadas y conservar los registros relevantes. Si el sistema tuvo una influencia

material, debe valorarse su comunicación a las partes y la incorporación de la documentación necesaria para el recurso.

Existe una prueba sencilla: si, al eliminar la salida de la IA, no somos capaces de explicar de forma autónoma por qué decidimos, la herramienta ha dejado de asistir y ha comenzado a sustituir.

10. CONCLUSIÓN: CONSERVAR LA RESPONSABILIDAD SIGNIFICA CONSERVAR EL JUICIO

La IA puede contribuir a una Justicia mejor. Puede reducir tareas repetitivas, facilitar acceso a jurisprudencia, ordenar información y mejorar claridad. Negar esa posibilidad sería tan poco prudente como aceptar que toda innovación es progreso.

Los tres riesgos examinados están conectados. El sesgo decide silenciosamente quién soporta los errores. La opacidad impide descubrirlos y discutirlos. La delegación indebida transforma una recomendación no comprendida en decisión revestida de autoridad judicial. Cuando coinciden, la firma humana puede convertirse en el último eslabón de una cadena gobernada fuera del proceso.

Las garantías jurisdiccionales no son obstáculos burocráticos. Son la tecnología constitucional que permite decidir legítimamente en condiciones de incertidumbre. Contradicción, motivación, independencia, individualización y recurso son mecanismos de control de errores. La IA debe integrarse dentro de ellos, no colocarse por encima ni al margen.

El futuro de la Justicia no será humano o algorítmico. Será una Justicia humana que utilizará algoritmos. Precisamente por eso debemos decidir qué puede hacer la máquina, qué debe seguir haciendo personalmente el juez y qué derechos corresponden a quien resulta afectado.

No necesitamos una inteligencia artificial que parezca juez. Necesitamos herramientas que ayuden a leer mejor, buscar con mayor precisión y explicar con más claridad, sin sustituir el encuentro responsable entre el Derecho y el caso concreto.

La pregunta final no es cuánto sabe la máquina. Es cuánto control jurídico, democrático y humano estamos dispuestos a conservar sobre ella.

REFERENCIAS NORMATIVAS, JURISPRUDENCIALES Y BIBLIOGRÁFICAS

Normativa y documentos institucionales

[Reglamento \(UE\) 2024/1689, de Inteligencia Artificial](#), en especial arts. 6, 9, 10, 12 a 14, 26, 27, 49, 85 y 86, y anexo III.

Reglamento (UE) 2026/1744 del Parlamento Europeo y del Consejo, de 8 de julio de 2026, por el que se modifican los Reglamentos (UE) 2024/1689, (UE) 2018/1139 y (UE) 2023/1230 en lo que respecta a la simplificación de la aplicación de normas

armonizadas en materia de inteligencia artificial (Ómnibus digital sobre IA), DO L, 2026/1744, de 24 de julio de 2026.

[Carta de los Derechos Fundamentales de la Unión Europea](#), arts. 7, 8, 20, 21, 47 y 48.

[Convenio Europeo de Derechos Humanos](#), arts. 6, 8, 13 y 14.

Consejo General del Poder Judicial, [Instrucción 2/2026, de 28 de enero, sobre la utilización de sistemas de IA en el ejercicio de la actividad jurisdiccional](#), BOE núm. 27, de 30 de enero de 2026.

CTEAJE, [Política de uso de la Inteligencia Artificial en la Administración de Justicia](#), versión 1.0, junio de 2024.

CEPEJ, [European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment](#), 2018.

Consejo de Europa, [Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law](#) (CETS n.º 225), 2024.

Jurisprudencia y decisiones

Tribunal Supremo, Sala Tercera, STS 1119/2025, de 11 de septiembre, rec. 7878/2024, ECLI:ES:TS:2025:3826, caso *BOSCO*.

TJUE, sentencia de 27 de febrero de 2025, *Dun & Bradstreet Austria*, C-203/22, ECLI:EU:C:2025:117.

TJUE, sentencia de 7 de diciembre de 2023, *SCHUFA Holding (Scoring)*, C-634/21, ECLI:EU:C:2023:957.

TJUE, sentencia de 27 de febrero de 2018, *Associação Sindical dos Juizes Portugueses*, C-64/16, ECLI:EU:C:2018:117.

TEDH (Gran Sala), *Baka c. Hungría*, demanda 20261/12, 23 de junio de 2016.

TEDH (Gran Sala), *Guðmundur Andri Ástráðsson c. Islandia*, demanda 26374/18, 1 de diciembre de 2020.

TEDH (Gran Sala), *Big Brother Watch y otros c. Reino Unido*, demandas 58170/13, 62322/14 y 24960/15, 25 de mayo de 2021.

TEDH, *Glukhin c. Rusia*, demanda 11519/20, 4 de julio de 2023.

Wisconsin Supreme Court, *State v. Loomis*, 2016 WI 68, 881 N.W.2d 749.

BIBLIOGRAFÍA

Alexy, R., *Teoría de la argumentación jurídica*, Centro de Estudios Constitucionales, Madrid, 1989.

Barocas, S.; Hardt, M.; Narayanan, A., *Fairness and Machine Learning*, 2023, [edición abierta](#).

Burrell, J., «How the Machine “Thinks”: Understanding Opacity in Machine Learning Algorithms», *Big Data & Society*, vol. 3, 2016, [DOI 10.1177/2053951715622512](#).

Chouldechova, A., «Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments», *Big Data*, vol. 5, n.º 2, 2017, pp. 153-163, [DOI 10.1089/big.2016.0047](#).

Citron, D. K.; Pasquale, F., «The Scored Society: Due Process for Automated Predictions», *Washington Law Review*, vol. 89, 2014, pp. 1-33, [texto abierto](#).

Cotino Hueso, L., «El uso jurisdiccional de la inteligencia artificial: habilitación legal, garantías necesarias y la supervisión por el CGPJ», *Actualidad Jurídica Iberoamericana*, n.º 21, 2024, pp. 494-527, [texto completo en la revista](#).

Dahl, M.; Magesh, V.; Suzgun, M.; Ho, D. E., «Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models», *Journal of Legal Analysis*, vol. 16, 2024, pp. 64-93, [DOI 10.1093/jla/liae003](#).

Dressel, J.; Farid, H., «The Accuracy, Fairness, and Limits of Predicting Recidivism», *Science Advances*, vol. 4, n.º 1, 2018, eao5580, [DOI 10.1126/sciadv.aao5580](#).

Dworkin, R., *Law's Empire*, Harvard University Press, Cambridge (Mass.), 1986.

Eubanks, V., *Automating Inequality*, St. Martin's Press, Nueva York, 2018.

Floridi, L. et al., «AI4People—An Ethical Framework for a Good AI Society», *Minds and Machines*, vol. 28, 2018, pp. 689-707, [DOI 10.1007/s11023-018-9482-5](#).

Green, B.; Chen, Y., «Disparate Interactions: An Algorithm-in-the-Loop Analysis of Fairness in Risk Assessments», *FAT** 2019, pp. 90-99, [DOI 10.1145/3287560.3287563](#).

Habermas, J., *El futuro de la naturaleza humana. ¿Hacia una eugenesia liberal?*, trad. R. S. Carbó, Paidós, Barcelona, 2002.

Hildebrandt, M., *Smart Technologies and the End(s) of Law*, Edward Elgar, Cheltenham, 2015.

Imai, K. et al., «Experimental Evaluation of Algorithm-Assisted Human Decision-Making», *Journal of the Royal Statistical Society: Series A*, vol. 186, n.º 2, 2023, pp. 167-189, [DOI 10.1093/jrsssa/qnad010](#).

Kleinberg, J.; Mullainathan, S.; Raghavan, M., «Inherent Trade-Offs in the Fair Determination of Risk Scores», *ITCS 2017*, art. 43, [DOI 10.4230/LIPIcs.ITCS.2017.43](#).

Lessig, L., *Code: Version 2.0*, Basic Books, Nueva York, 2006, [edición abierta](#).

Llano Alonso, F. H., *El jurista ante los retos de la revolución tecnológica 4.0*, en A. M. Marcos del Cano (coord.), *Inteligencia artificial. Desafíos éticos, sociales y jurídicos*, Dykinson, Madrid, 2026, pp. 21-36.

Marcos del Cano, A. M. (coord.), *Inteligencia artificial. Desafíos éticos, sociales y jurídicos*, Dykinson, Madrid, 2026, 172 pp., ISBN 979-13-7047-178-1, [ficha bibliográfica e índice](#).

O'Neil, C., *Weapons of Math Destruction*, Crown, Nueva York, 2016.

Ost, F., *Júpiter, Hércules, Hermes: tres modelos de juez*, Academia, 2007.

Pasquale, F., *The Black Box Society*, Harvard University Press, Cambridge (Mass.), 2015.

H. Suresh; J. V. Gutttag, *A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle*, Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO 2021), ACM, 2021, art. 17, pp. 1-9, [DOI 10.1145/3465416.3483305](#).

Sotomayor Padilla, P. C. et al., «Jurisprudencia sobre inteligencia artificial y responsabilidad jurídica», *Ethos Scientific Journal*, vol. 4, n.º 1, 2026, pp. 147-165, [DOI 10.63380/esj.v4n1.2026.301](#).

Taruffo, M., *La prueba de los hechos*, Trotta, Madrid, 2002.

Inteligencia Artificial generativa

Panorama desde un escritorio

Miguel Ángel Serrano

Escritor, experto en comunicación y presidente de honor de European Writers' Council.

Ha publicado un ensayo crítico sobre IA y cultura:

Androiceno. Escribir en la era de la inteligencia artificial. <https://n9.cl/m1kv2>

Lo que ya se puede testimoniar sobre la Inteligencia Artificial generativa

Es un hecho conocido que los Modelos Grandes de Lenguaje, base de los sistemas de inteligencia artificial generativa (la que, en lenguaje natural, ofrece las aproximaciones estadísticas más probables a las peticiones que se le hacen, en textos, imágenes o música) han sido y siguen siendo entrenados con materiales sujetos a *copyright* o, yendo más allá, que ponen en entredicho la legislación sobre protección de datos. Aunque esto afecta a todos los ciudadanos (especialmente por el uso de redes sociales como Facebook o Twitter, que pertenecen a compañías que también desarrollan modelos de inteligencia artificial) los sectores creativos están viendo cómo sus obras sirven para alimentar modelos que compiten directamente contra ellos: los escritores hemos de ver cómo se genera un flujo inacabable de novelas o ensayos “cometidos” mediante inteligencia artificial, los ilustradores pierden mercado a gran velocidad, los traductores ven bajadas de ingresos o de su propia consideración profesional. Es algo que debería ponernos en alerta.

Me he ocupado por extenso de este asunto en mi libro *Androiceno: escribir en la era de la inteligencia artificial* (Berenice, 2025), en el que analizo las intersecciones entre esta tecnología y el sector cultural, aunque con una mirada totalizadora que también comprende aspectos sociales, económicos y laborales. Como expreso en el volumen, las consecuencias del despliegue incontrolado de la tecnología, que parece más una operación militar, son devastadoras en algunos campos, como lamentablemente estamos empezando a ver. Por ejemplo, y de ahí el título del libro, el impacto medioambiental es desaforado: “Androiceno” es el término que propongo para explicar que la introducción de agentes autónomos implica

que las máquinas toman decisiones para su prevalencia (y ya hay ejemplos de esto) sin tener en cuenta otros factores, como el consumo de agua. Sería la superación del término Antropoceno, esa teoría que afirma que el destino del planeta está ya solo en nuestras manos. Desde ahora, también en las manos de las máquinas o como mínimo, del oligopolio que está absorbiendo recursos para acelerar el desarrollo y despliegue.

Por otro lado, desde 2023 hasta el pasado año, presidí el Consejo Europeo de Escritores justo en la eclosión pública de la tecnología, lo que me situó en primera línea del debate político en pos de una regulación que consiga proteger a creadores y consumidores: una batalla en múltiples frentes, desde la evidente necesidad de pagar (y pedir permiso previo) por el uso de contenidos sujetos a cortapisas de propiedad intelectual (incluidos los publicados en internet) a una información veraz y comprobable por el consumidor. Del mismo modo que exigimos etiquetados que aseguren que el yogur es yogur, deberíamos exigir que un texto o imagen producidos con inteligencia artificial puedan ser identificados, o que las pseudo-obras artísticas no se vean beneficiadas por reducciones de IVA o reciban subvenciones destinadas a obras o empresas culturales. Del mismo modo, los productos generados con inteligencia artificial no son susceptibles de generar *copyright*, y tampoco deben ser autorizados a participar en concursos literarios que no lo especifiquen, y desde luego no los pagados con dinero público.

Todo ha cambiado

La ola no cesa. Afecta y afectará a todos los sectores: desde los gobiernos a la judicatura, desde la consultoría a las series de televisión. Los mercados tratan de ajustarse ante un paisaje legal que es distinto dependiendo de la zona en la que se ponga

el foco: se acusa a la Unión Europea de sobre regular, pero por otro lado vemos una fuerte renuencia en Estados Unidos (que está empezando a tener diferencias por las legislaciones de los Estados) para dejar expedito el camino, a cambio de atropellar los derechos de millones de personas, a las grandes multinacionales norteamericanas del sector, que están inmersas en una carrera alocada con empresas chinas y en la que estas parecen reducir la distancia a gran velocidad, lo que plantea un escenario de conflicto. Ya se oyen rumores sobre la prohibición de modelos chinos en Estados Unidos. En realidad, si se considera desde el punto de vista geopolítico, podríamos decir que estamos en una guerra de sexta generación, cuyo botín es la captura de la cognitividad de las poblaciones, intención que ha declarado abiertamente el gobierno chino: no hay que olvidar que estos sistemas están plagados de sesgos que pueden desembocar fácilmente en armamento propagandístico. No hay que desdeñar, con todo, el efecto Bruselas: incluso con las graves deficiencias de la ley, abrió el camino a otras zonas, como Australia o la India, para que el asunto empezase tomarse en serio por los legisladores. Incluso China está generando normativa activamente y con una mirada avanzada.

Ahora bien, la ausencia de regulación efectiva está mostrando la potencia y el peligro de la tecnología (incluso desde sus paupérrimos fundamentos éticos) en áreas que suponen problemas serios. Por ejemplo, lo que se conoce como descarga cognitiva, esto es, confiar de modo regular y acrítico en las respuestas que dan los diversos modelos, lo que no es sino una manera de eliminar la fructífera actividad intelectual que supone pensar (y escribir) por uno mismo. Más grave aún parece ser la relación de dependencia y autoafirmación que los programas suponen para poblaciones en riesgo emocional o psicológico, como pueden ser niños y adolescentes o personas que pueden verse envueltas en una escalada psicótica. Hemos visto suicidios inducidos en la infancia. De manera menos trágica, y en el campo legal, hay noticia de los errores que los modelos cometen en, por ejemplo, la inclusión de jurisprudencia falsa en instancias judiciales, lo que supone un grave riesgo para todo el sistema. Hay ejemplos que, de no implicar peligrosas consecuencias, resultarían cómicos. En el caso de profesiones colegiadas, como psiquiatras o abogados, la residencia de la responsabilidad no está clara, aunque estamos viendo como las grandes tecnológicas intentan no asumir ninguna por el uso de sistemas que, claramente, son defectuosos cuando no viciados desde el diseño. Es todo un reto imaginar otro sector en el que esto se permita. Y, como aviso a navegantes, son los usuarios quienes finalmente deben hacerse cargo de las consecuencias, como hemos visto en el caso de grandes consultoras que se han visto obligadas a devolver parte de lo facturado por usar inteligencia artificial de forma, digamos, poco profesional. Esta tecnología está generando un serio problema de cumplimiento para muchas empresas.

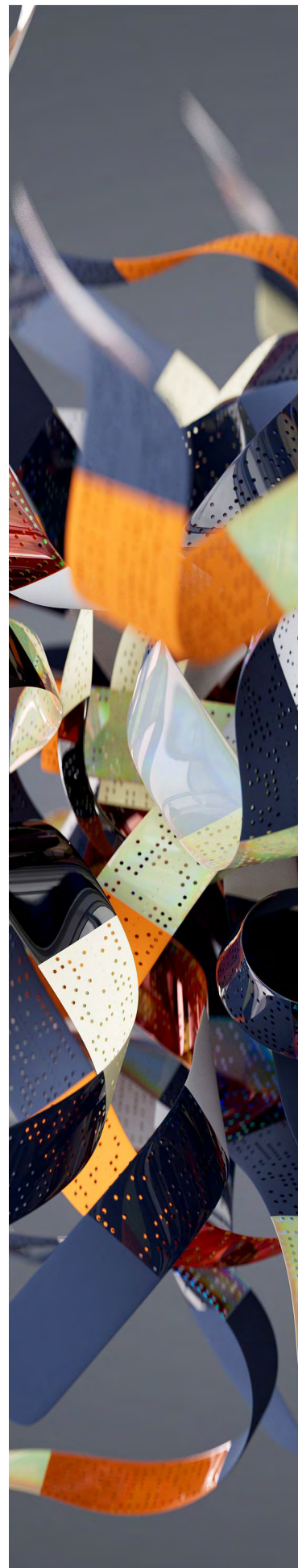
Naturalmente, la litigiosidad aumenta en varios países, especialmente Estados Unidos. Por un lado, familiares de personas lastimadas por el uso descontrolado del sistema (por ejemplo, adolescentes suicidas), empresas que usan agentes de IA que proporcionan información incorrecta, etc. Por el otro, multitud de artistas, editoriales, productoras o todo tipo de detentadores de derechos de autor o personas cuyos datos han sido utilizados sin permiso, por no hablar de los *deepfakes*, que lesionan gravemente la reputación. Ya estamos viendo cómo agentes de inteligencia artificial han provocado destrozos en las bases de datos de las empresas usuarias, y el catálogo de posibilidades de error no hace más que aumentar. Eso, sin entrar en las autopistas enormes que estos sistemas abren para todo tipo de delincuentes.

Internet es parte del problema

Algo llamativo es que las grandes tecnológicas (GAFAM: Google, Apple, Facebook, Amazon y Microsoft) cuentan con modelos de inteligencia artificial o se sitúan, como Amazon, en la cadena de valor de los centros de computación. Precisamente el crecimiento esperado de este tipo de granjas de computación es uno de los grandes problemas medioambientales del futuro inmediato: inversiones inabarcables para mantener la temperatura de mega centros que apenas generan empleo y que suelen empeorar la calidad medioambiental (en especial el agua) de las zonas en las que se instalan. Aragón, por ejemplo, cuenta con un ambicioso plan de inversiones de este tipo, y los ciudadanos empiezan a preguntarse sobre la pertinencia de usar tantos recursos hídricos para refrigerar máquinas. Se habla ya de montar granjas en el espacio, que cuenta con la ventaja de la refrigeración gratuita.

¿Por qué es internet parte del problema? Si unimos a Elon Musk al repóker de oligopolistas veremos que el dominio digital, incluso con la concurrencia de algunos jugadores chinos como Baidú o Alibabá, es estable desde hace años. De hecho, internet ha ido generando, a través del uso interesado de las licencias *Creative Commons*, que relajan las exigencias del *copyright* y que han sido fuertemente patrocinadas por estas empresas, una suerte de idea generalizada de que lo que circula por internet es público. Esto implica confundir lo publicado y lo público, aunque la confusión se acaba en el momento en el que intentemos obtener de manera gratuita un programa de ordenador de Microsoft, por ejemplo. Y, sin embargo, todo internet ha sido rastrillado, aspirado, troceado y usado en el entrenamiento de los Modelos Grandes de Lenguaje. Incluso modelos generados con dinero público español han utilizado repositorios como Common Crawl, que, digamos, no parecen sostener una correcta posición ética. Esto está en la base de la lucha de los autores por sus derechos, como ampliaremos luego.

A menudo me encuentro con personas que sostienen este argumento. Incluso altos oficiales del sector de la IA han confundido la accesibilidad con el permiso al usar, por ejemplo, todo lo que hay en YouTube. Siempre pongo el mismo ejemplo: que las frutas de mis árboles estén al alcance la mano no significa que todo el mundo pueda cogerlas. O, si me encuentro una cartera llena de billetes, lo lógico es llevarla a una comisaría, esperar el plazo legal para la reclamación y entonces hablar con Hacienda. Quedármela es ilegal, llanamente.



Al usuario le suele hacer menos gracia enterarse de que todo lo que haya subido a una red social ya ha sido devorado para entrenar a la inteligencia artificial, y que incluso puede ser parte de programas de vigilancia y reconocimiento facial, cuando no de videos *fake* ilegales y ofensivos. Nadie está libre de esto, y lo peor es que cualquiera puede hacer un video estúpido de, por ejemplo, Chuck Norris apalizando a un político. Mientras, el avance imparable de la red está causando problemas estructurales en las ciudades, de las que desaparece el pequeño comercio local, y con él la satisfacción de impuestos granulares, lo que ya está dificultando la financiación de, por ejemplo, los sistemas de transporte público. Además, las calles pierden vida y proliferan los locales reconvertidos a vi-



“Muchos de los problemas que asociamos a la inteligencia artificial se replican o tienen su base en las redes sociales, las grandes formadoras de opinión de hoy en día y que están en muy pocas manos”

viendas de baja calidad. Y esto alcanza a ciudades del primer mundo, como Nueva York, que se enfrenta a graves dificultades de financiación.

Muchos de los problemas que asociamos a la inteligencia artificial se replican o tienen su base en las redes sociales, las grandes formadoras de opinión de hoy en día y que están en muy pocas manos. Hemos visto ya ejemplos de propaganda rusa o china y la perpetuación de sesgos o prejuicios. No en vano se dice que la inteligencia artificial es *WEIRD* (rara): *Western, Educated, Industrialised, Rich, Democratic*. Yo me permito poner en duda lo último: la concentración de riqueza en manos de menos personas y las crecientes desigualdades sociales y nacionales plantean algunas cuestiones que convendría resolver o al menos debatir en serio. Hace tiempo que está claro que la inteligencia artificial es más política que tecnología.

¿Mantequilla en las dos caras de la tostada?

Si algo puede salir mal, saldrá mal: esta es la máxima del pesimista, pero también de los analistas de riesgos. Confía, pero verifica: esta es la máxima del optimista informado. Corre y rompe cosas: esta es la máxima de las grandes tecnológicas, que utilizan su gran poder para salir indemnes de esas roturas. Algunos podrían decir que eso es optimismo, yo creo que es sensación de impunidad.

El debate sobre la IA, en realidad, es una disputa sobre la preeminencia del relato. Por un lado, existe un gigantesco esfuerzo de marketing que trata de convencernos no solo de las bondades de la tecnología (que algunas tiene, por descontado), sino de quitarle hierro a las maldades (que también tiene, claro). El debate se extiende a las personas: hemos llegado a leer cómo a

quienes se oponen (nos oponemos) al uso sin regulación de la tecnología se nos llama anticristos, nada menos. El objetivo es forzar la sensación de inevitabilidad de un advenimiento de lo que se da en llamar inteligencia artificial general, que implicaría que los sistemas informáticos nos superan en cualquier tarea cognitiva, algo que parece difícil según el estado actual de la tecnología.

El relato, naturalmente, opera a muchos niveles, con una legión de partidarios en gran medida acríticos que constantemente menosprecian a los que no lo son: “vas a perder el trabajo si no te haces experto en IA”, “eres un retrógrado, un neoludita”, “la regulación impide la innovación”, “se van a crear más empleos de los que se van a destruir”... afirmaciones difícilmente verificables y basadas generalmente en los deseos y en un gigantesco sesgo de efecto arrastre que es casi imposible de rebatir. Y esto es así porque a muchos usuarios la tecnología les proporciona una ventaja real: gente que no sabe escribir publica “obras” (“comete contenidos” sería más adecuado), pseudo profesionales cobran por consultar sin conocimientos valiosos, o se ejecuta código sin tener una formación sólida en ello. Eso suele llevar aparejada una férrea defensa de actos no del todo éticos, o no siempre: imagine el lector qué no dirán los directivos de las empresas tecnológicas...

En el otro lado, los apocalípticos predicen la llegada de una superinteligencia que esclavizará y destruirá a la raza humana. Lo curioso es que este discurso es también adoptado por los mercadólogos de las tecnológicas, empeñados en decir que la inteligencia artificial ya está aquí, que las máquinas son seres sintientes, que algunos sistemas son demasiado poderosos para lanzarlos al mercado o que, y este es el asunto, las máquinas tienen derechos, algo que debería horrorizar a cualquier jurista, pero que implicaría una enorme ventaja para las empresas. En primer lugar, porque todos esos textos producidos mediante “máquinas de plagio” como las llama Chomski, podrían producir derechos de autor, algo que actualmente no es posible y que limita mucho, como debe ser, las posibilidades comerciales de, por ejemplo, las películas producidas mediante esta tecnología.

Bienvenidos a la demorragia

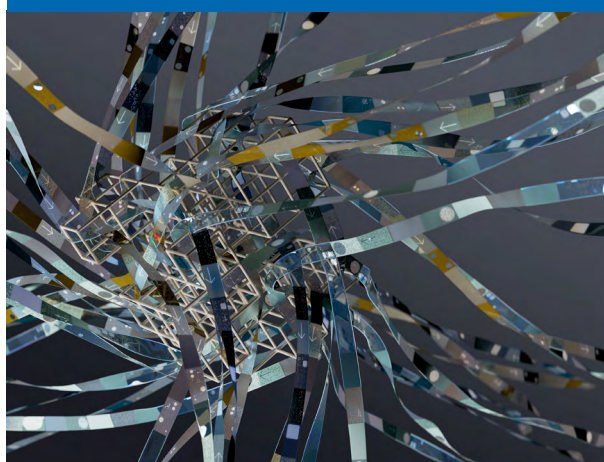
Que los sistemas democráticos están en peligro no parece una noticia. Cualquiera que vea la progresión de las sociedades donde este sistema opera con cierta calidad, y no son tantas, observará con preocupación el populismo (fuertemente alimentado por las redes sociales), la erosión de los mecanismos de contrapeso institucional y *check and balance* y el creciente control social por parte de los gobiernos. La inteligencia artificial es combustible para estos problemas. Por ejemplo, la polarización es fuertemente alimentada por *bots* (al punto que se calcula que la mitad del tráfico en internet es generado por máquinas), los medios se ven abocados a una enorme pérdida de ingresos, los periodistas, como los ilustradores o los escritores, se ven condenados a un futuro de bajos emolumentos, la tentación de la automatización en determinadas partes del sistema judicial (pero también legislativo, con la idea descabellada de sustituir los parlamentos por sistemas de IA más eficientes o votaciones plebiscitarias constantes) o, en fin, el control asfixiante del comportamiento de los ciudadanos, que permite no solo la ubicación, sino el perfilado psicosocial de precisión cuando no la injerencia directa, mediante la moneda digital centralizada, en hábitos o gastos so capa de mejorar la eficiencia de los sistemas de salud, etc.

A esto lo denomino demorragia, y el proceso tiende a acelerar. Pensemos en el caso de la ministra virtual de Albania, Die-lla. Un sistema de inteligencia artificial que nace para mejorar la contratación pública y evitar la corrupción. El gobierno la nombra oficialmente ministra, (lo cual parece inconstitucional, según reclama la oposición), la publicita, le dota de un discurso y finalmente, las dos profesionales que la crearon son acusadas de... corrupción. Al fin y al cabo, si el trabajo de un programador, el de un redactor o el de un traductor pueden ser sustituidos por un programa, ¿por qué no el de una ministra? La pregunta es evidentemente tramposa, pero a muchos parece no importarles.

La actividad del poder legislativo tampoco está exenta del peligro de la automatización. Emiratos Árabes Unidos está utilizando ya activamente esta tecnología para actualizar y producir gran parte de su corpus legislativo, y probablemente esta es la primera de muchas actuaciones de este tipo en otros países. El problema de esto es que el debate, la fricción, la consulta, son elementos constitutivos de la producción de leyes, incluso aunque a veces el proceso sea exasperantemente lento. El viejo aserto alemán de que es mejor no saber cómo se hacen las salchichas ni las leyes parece no ser tan evidente o sabio. Porque la opacidad algorítmica es un hecho, y la falta de reproducibilidad de la iteración de la producción de respuestas de la IA también.

Por último, ¿estará el poder judicial entredicho? Por supuesto. En cualquiera de los casos no parece una gran idea entregar a una tecnología que tiene dueño y que ni siquiera es nacional o comunitaria los poderes del Estado. Siempre se aduce que la obligatoriedad del control humano, como en el caso de las armas susceptibles de funcionar autónomamente, va a ser protegido y garantizado, pero habría que preguntarse por cuánto tiempo. Por ejemplo, muchas controversias online, o

“Vista la velocidad del desarrollo de la tecnología, no es irrazonable pensar que muchas de las funciones de jueces o letrados pueden ser sustituidas a corto plazo”



las que se den en determinados servicios de atención al cliente, son ya solucionadas por algoritmos y es inevitable pensar que es un enorme campo de pruebas. Vista la velocidad del desarrollo de la tecnología, no es irrazonable pensar que muchas de las funciones de jueces o letrados pueden ser sustituidas a corto plazo.

No solo eso: la potencia de algunos programas es tal que hay cierto temor entre los gobiernos a quedarse atrás. Si admitimos que la IA es también política, es difícil sustraerse a la idea de que pueda usarse con fines sometidos a estrategias nacionales. Un ejemplo sería la petición de la Comisión Europea, y la preocupación de los estados miembro, ante el lanzamiento de sistemas poderosos, como Mythos, que promete detectar vulnerabilidades de ciberseguridad, al parecer con excelen-

tes resultados y que ha sido distribuido a unos pocos actores, entre otros, bancos estadounidenses que pueden así mejorar su arquitectura de seguridad, mientras los grandes bancos europeos pueden verse amenazados y superados. De ahí que la Comisión presionara para que se incluyera a actores europeos en la distribución limitada del sistema. Los acontecimientos posteriores, con la prohibición de uso de Fable 5 para ciudadanos no estadounidenses, finalmente retirada, hizo que muchos se dieran cuenta de que estamos en un mercado de licencias que puede ser manipulado por el gobierno norteamericano (y el chino, que es un gobierno dictatorial). Esto ha encendido el debate sobre la soberanía tecnológica europea.

Un lado de la tostada dice adoptémosla rápidamente, el otro dice no lo hagamos. Y en este caso, parece difícil ponerse de lado.

Es la economía, estúpido

La famosa interpelación de Bill Clinton a su oponente es en realidad un componente imprescindible de cualquier análisis sobre inteligencia artificial: una tecnología tan disruptiva como esta, que se está desplegando a enorme velocidad y en todo el mundo, supone un sector económico de enorme magnitud. La economía que sostiene una corriente de inversión inacabable en empresas que pierden dinero de manera escandalosa, como Open AI, está en entredicho, y muchos consideran que estamos ante una burbuja. Hay que tener en cuenta que la rentabilidad (esperada y prometida, no real) se basa en una adopción masiva, especialmente por parte de las empresas, que no se está dando. Las versiones gratuitas de los programas más populares son las que realmente se usan, así como la curación de búsquedas por internet por parte de estos sistemas, lo que además cambia drásticamente el paisaje económico de la publicidad en internet, con consecuencias onerosas para medios y anunciantes.

No obstante, lo que a veces se conoce como *uberización*, esto es, la sustitución de un sector tradicional (el taxi, en ese caso) por otro apoyado por la tecnología y con precios dinámicos, capaces de ajustarse a la demanda, suele empezar con servicios muy baratos o gratuitos. Sin embargo, el mandato, perfectamente legal y ético, vaya por delante, de buscar rentabilidad, hace que la estructura de precios varíe cuando se alcanza una posición preponderante. La inteligencia artificial generativa a bajo coste o gratuita no parece sostenible, al menos no para las empresas desarrolladoras. La estrategia es la típica del caramelo (o de las drogas): consigo que te descargues cognitivamente en la máquina (esto es, que llegues a pensar que no sabes escribir un email, por ejemplo) y luego empiezo a cobrarte. En realidad, la operación de privatización del conocimiento humano (que es lo que significa, en última instancia) implica que en algún momento vas a tener que pagar: de nuevo la narrativa marketiniana se pondrá en marcha para limar la disonancia cognitiva. Existe la televisión gratuita y existe Netflix, por poner un símil. Pero si no pagas, algo malo te va a pasar: tu empresa cerrará, te quedarás sin trabajo, etc.

¿Cómo sostener los enormes costes de computación y hardware de estos sistemas? Hay intentos de generar sistemas de propiedad pública, como el caso de MarIA, sufragado por el gobierno español, y más ejemplos en otros países de Europa. La promesa de muchas empresas chinas es que trabajarán en abierto, del mismo modo que

lo proclama Llama, propiedad de Facebook. Si hasta el perejil se cobra, yo no me fiaría mucho de estas promesas de gratuidad eterna. La otra falacia es el bajo coste. En este momento el castigo lo sufren los inversores, muchos de ellos atrapados en la trampa del *too big to fail*, pero la tensión económica se trasladará tarde o temprano a los usuarios, como de hecho empieza a ocurrir. Algunas empresas se encuentran facturas por el uso de agentes de inteligencia artificial que son más gravosas que las contrataciones de personas. Si se paga por consumo, el escenario cambia, y empezamos a ver cómo la codificación mediante IA es bastante más cara de lo que se esperaba. Tener cuatro o cinco agentes de IA al servicio de un empleado puede resultar una inversión ruinosa si se paga por uso.

El sector de la inteligencia artificial impacta en muchos otros. Empezamos a ver despidos masivos en todo el mundo bajo la bandera “*AI first*”, que en mi opinión debería bastar para dejar de usar los servicios o consumir los productos de las empresas que proclamen tal cosa. En realidad, muchos analistas consideran que la IA es una excelente excusa para este tipo de movimientos sin que en realidad haya una relación directa. Consultoras, banca, servicios de atención al cliente y otros sectores van a sufrir seriamente. Otra derivada es qué ocurrirá con el consumo. Si, como se augura por los propagandistas, los trabajos mejor pagados (seniors, directivos, arquitectos, etc.) van a ser sustituidos, ¿quién sostendrá el mercado inmobiliario? ¿Las pensiones? Una solución que se apunta es la renta universal, pero parece un oxímoron. ¿De verdad alguien piensa que esa renta va a llegar a los países del tercer mundo? En realidad, según la tecnología se perfecciona, muchos trabajos deslocalizados en países en desarrollo van a desaparecer: adiós a los *call center* en países asiáticos o hispanoamericanos...

En el principio de esta cadena está el desvío evidente de renta que se trasvasa desde sectores que gozan (o gozaban) de buena salud y proporcionan millones de puestos de trabajo, como las industrias creativas o las informáticas, cuya potencia económica se debilita, como el roble frente al muérdago, mediante la sustitución de su servicio o producto por elementos producidos con IA. El cine, los actores, guionistas, escritores, traductores, programadores, diseñadores, fotógrafos e ilustradores... cientos de profesiones están abocadas a la desaparición o como poco a la pauperización.



Inteligencias de artificio

Ars longa, vita brevis. La famosa máxima de Hipócrates podría traducirse por “El arte es largo (costoso) de aprender, la vida es breve”. En realidad, la frase completa rezaría “La vida es breve, el aprendizaje, largo; la ocasión, fugaz; la experiencia, insegura; el juicio, difícil”. Si algo me parece peligroso es la promesa de eliminar la fricción, el esfuerzo, para conseguir algo. Un doctorado debe ser difícil, el verdadero juicio, provisional y abierto a la mejora, la experiencia, un camino que no necesariamente proporciona aprendizaje, sino tan solo achaques. Y la ocasión fugaz (la inspiración, por ejemplo) deja de serlo cuando se puede obtener bajo demanda.

Llevo mucho tiempo defendiendo que escribir es pensar, que labrar un destino de artista significa sobre todo iteración, trabajo, autocrítica. Como he reivindicado en algunos foros de la UE, tenemos que preservar al derecho a ser artistas. La apuesta vital, muchas veces ardua, no debería verse cercenada por una tecnología que elimina el esfuerzo. Multitud de profesores comprenden con angustia que cada vez más lo que hacen es poner nota a la habilidad para usar ChatGPT, los editores viven con la amenaza de que les cuelen una “novela” “cometida” con inteligencia artificial (ya ha habido casos de rupturas de contrato por este motivo), muchos escritores reclaman un sello que certifique que algo está hecho sin IA. Yo abogo por lo contrario: es esta tecnología la que está usurpando la legitimidad autoral de la obra. Y también apuesto por fuertes castigos ante posibles incumplimientos. Es necesario repensar la estructura de la propiedad intelectual, pero no para debilitarla, como reclaman las empresas desarrolladoras, sino para establecer con claridad que es una propiedad positiva, que no necesita afirmación una vez establecida. Y está protegida en la Declaración Universal de los Derechos Humanos, concretamente en el artículo 27,

además de en el 17 de la Carta de Derechos Fundamentales de la Unión Europea. Yo he introducido en mi libro de idea de los Derechos de los Humanos, que implica la defensa de las personas frente a la presencia impuesta y la búsqueda de derechos para las máquinas: por ejemplo, reivindicar el derecho a no ser obligado a utilizar esta tecnología o a no sufrir consecuencias por no hacerlo, a no ser juzgado por un robot en litigios de calado, a no ceder el control de la autonomía monetaria o a no sufrir presiones para abrazar ideas transhumanistas.

Lo que en realidad debería preocuparnos es que, si no conseguimos generar un mercado estable, y rentable, para las profesiones creativas, y permitimos que la inteligencia artificial siga utilizando materiales protegidos para crear un mercado alternativo que compite directamente con las obras futuras de los detentadores de derechos, en una sangrante pirueta, no habrá especiales incentivos para la creación. Si ya es un tipo de profesión depauperada y sometida a una enorme inestabilidad, la erosión del margen puede matar la cultura. Esta protección solo puede conseguirse con regulación. Fortaleciendo la propiedad intelectual y evitando atajos como la directiva DSM (2019/790) a cuya excepción del artículo 3, que permite el uso de obras protegidas sin permiso para la investigación sin ánimo de lucro, se agarran los desarrolladores para justificar un expolio que algunos consideran el mayor robo de propiedad intelectual de la historia. Yo entre ellos.

Estamos matando al siguiente Kafka, Picasso, Georgia O’keefee. Y lo estamos permitiendo sin apenas pestañear, cegados por las pantallas de nuestros teléfonos. Este es el panorama que se ve desde mi escritorio, pero no quiero refugiarme en mi obra, mirando mi ordenador con la cabeza gacha y sin observar y criticar lo que me rodea. Seguir escribiendo sobre ello es una decisión ética. Eso que no tienen las máquinas.