Revista Crítica
de
Relaciones de Trabajo

Laborum

Estudios de Doctrina Judicial
nº 20 (3er Trimestre 2026)
ISSN: 2792-7962 – ISSNe: 2792-7970
Págs. 155-175

¿Serán sostenibles los principios del proceso social ante el uso negligente de la IA? ¿Disponemos de vías garantistas de detección e impugnación ante una *praxis* viciada por el juzgador o la contraparte?

Will the principles of social process be sustainable face to a negligent use of AI? Do we have procedural safeguards to detect and contest vitiated practices by the court or the opposing party?

Ainhoa Nieto Garrote *Investigadora predoctoral*
Universidad de Deusto
https://orcid.org/0009-0009-7156-2936

Sumario

1. **Mala *praxis* emergente debida al sesgo de automatización**
2. **La necesidad de una alfabetización técnica**
   2.1. Diversidad taxonómica de los errores técnicos en el ámbito jurídico
   2.2. Sobre los sistemas RAG
   2.3. Problemas subyacentes que amenazan la integridad de los procesos judiciales en general
3. **Principales problemas detectados en la Instrucción 2/2026 del CGPJ, sobre la utilización de sistemas de IA en el ejercicio de la actividad jurisdiccional**
4. **Traducción de la inseguridad jurídica que provoca esta realidad en el proceso social**
   4.1. ¿Recurso de reposición? ¿Constituye una verdadera garantía teniendo en cuenta la celeridad del proceso?
   4.2. ¿Qué encaje y valoración tendría el sesgo de automatización en el juzgador si se detectase a tiempo para ulteriores recursos?
5. **Algunas propuestas acompañando a la depuración normativa planteada**
   5.1. Resoluciones judiciales con un apartado final donde se declare detalladamente si se ha empleado la IA para su formulación
   5.2. Mientras tanto, cabe solicitar el acceso al código fuente al amparo de la jurisprudencia del caso Bosco
   5.3. ¿Podría automatizarse un sistema pericial que revise toda la prueba obrante en autos junto a los escritos y las grabaciones de las vistas?
   5.4. ¿Los Ilustres Colegios de la Abogacía deberían contribuir a democratizar el acceso a herramientas de IA especializadas?
6. **Conclusiones**
7. **Bibliografía**

Fecha Recepción: 20/4/2026
Fecha Revisión: 7/6/2026
Fecha Aceptación: 8/7/2026

Cita Sugerida: NIETO GARROTE, A.: «¿Serán sostenibles los principios del proceso social ante el uso negligente de la IA? ¿Disponemos de vías garantistas de detección e impugnación ante una praxis viciada por el juzgador o la contraparte?». *Revista Crítica de Relaciones de Trabajo, Laborum,* 20 (2026): 155-175.

Resumen

Motivado por la mala *praxis* emergente en torno al empleo negligente de la IA debido al sesgo de automatización y las ultrafalsificaciones, mediante el presente artículo se sostiene la necesidad de reflexionar sobre la sostenibilidad de los principios vertebradores del proceso social y las garantías para detectar e impugnar los vicios que este arrastre. A su vez, se exponen los que se consideran puntos de mejora de la Instrucción 2/2026 del CGPJ y se proponen algunas medidas preventivas.

Palabras clave

Sesgo de automatización; alfabetización interdisciplinar; inseguridad jurídica; Instrucción 2/2026; medidas preventivas.

Abstract

Motivated by the emerging malpractice surrounding the negligent use of AI driven by automation bias and deepfakes, this article argues for the need to critically reflect on the sustainability of the core principles of the social process and the safeguards to detect and challenge any procedural flaws. Furthermore, it outlines potential areas for improvement regarding Instruction 2/2026 of the General Council of the Judiciary and proposes some preventive measures.

Keywords

Automation bias; interdisciplinary literacy; legal uncertainty; Instruction 2/2026; preventive measures.

1. MALA *PRAXIS* EMERGENTE DEBIDA AL SESGO DE AUTOMATIZACIÓN

El presente artículo surge a raíz de la imbricación, por un lado, de las dinámicas negligentes emergentes relacionadas con el uso de la IA en el ámbito judicial por parte de los operadores jurídicos, y, por otro lado, del cuestionamiento en torno a la salvaguarda de la seguridad jurídica conforme a los principios vertebradores del proceso social y algunos de sus hitos ordinarios.

Como consideración previa contextual, ¿estamos inmersos en un bucle de esqueumorfismo de los operadores jurídicos, quienes vician su actuación como consecuencia del sesgo de automatización?[1] El término esqueumórfico proviene del ámbito del diseño y puede definirse como el uso de características de diseño que hacen que un artefacto se parezca a una versión anterior de sí mismo[2]. Este concepto se ha aplicado al campo de la toma de decisiones semi automatizadas o plenamente automatizadas. Curiosamente, el diseño puede estar previsto para contener detalles como retrasos (innecesarios) que den a entender que existe una deliberación por parte del sistema. Mediante el trabajo que se referencia en este punto, la conceptualización descrita se aplica para analizar los tipos de participación humana en la toma de decisiones automatizadas en general[3], mientras que en el presente trabajo se traslada al ámbito judicial: ¿en qué medida estos operadores jurídicos se apoyan en los sistemas de IA para la toma de decisiones y cuál es la apariencia que se proyecta?

En el ámbito de la toma de decisiones y cómo se transmiten, las dos caras de la moneda quedan representadas como: de un lado, las dinámicas que presentan un ideal de automatización completa pero que, sin embargo, dependen de la actividad humana, y, *contrario sensu*, las dinámicas que sostienen un ideal de que el sistema de IA sea una herramienta de apoyo recayendo la actividad principal sobre el factor humano, pero que, al contrario, dependen del sistema de IA (es decir, la persona acaba careciendo de una capacidad significativa para influir en el proceso de toma de decisiones)[4]. Uno de los principales problemas derivados es que cuando la automatización es mayor a la que se aparenta, las personas afectadas por los resultados no tienen una oportunidad real de apelación.

Por ejemplo, para el primer supuesto, en el caso de las personas trabajadoras que se dedican a la moderación de contenidos (es una falsa automatización porque las plataformas de microtareas en línea tratan de dar la imagen de que la moderación de contenidos de las redes sociales es una tarea plenamente automatizada, cuando en realidad son las moderadoras de contenido las que se dedican a visualizar y decidir sobre su eliminación). Eso provoca una desalineación respecto a la posible impugnación de los juicios del sistema. Este ejemplo de falsa automatización interesa a las

[1] El sesgo de automatización representa la inclinación de los seres humanos a dar acríticamente por buenas las respuestas facilitadas por, en el caso que aquí interesa, la IA.

[2] «En el diseño esqueumórfico, lo que antes era funcional se vuelve ornamental, un guiño a la tecnología anterior que ayuda al usuario en la transición. Por ejemplo, el sonido del "clic del obturador" de la cámara de un teléfono: aunque la cámara ya no tiene un obturador físico que haga ese sonido, nos hemos acostumbrado a la idea de que el clic del obturador indica que se ha tomado una foto. Por lo tanto, las tecnologías posteriores han incorporado el sonido como un adorno. Este adorno conserva su función social al actuar como un significante, una notificación para quienes toman las fotos y para quienes las retratan de que se ha capturado una imagen. Pensemos también en los lectores electrónicos con "páginas" o en los controles de audio digital con forma de dial». Brenan-Márquez, K. Levy, K. Y Susser, D.: Strange Loops: Apparent versus actual human involvement in automated decision making. *Berkeley Technology Law Journal* vol. 34, 2019, p. 754.

[3] Brenan-Márquez, K. Levy, K. Y Susser, D.: Strange Loops: Apparent versus actual human involvement in automated decision making. *Berkeley Technology Law Journal* vol. 34, 2019. Pp 745–772.

[4] «En estos casos, el humano es, en la práctica, poco más que un elemento ornamental de la interfaz del sistema. Estas dinámicas surgen en sistemas técnicos o burocráticos que, aparentemente, involucran a humanos, pero donde estos son incapaces de ejercer discreción o apartarse de los guiones administrativos» Brenan-Márquez, K. Levy, K. Y Susser, D.: Strange Loops: Apparent versus actual human involvement in automated decision making. *Berkeley Technology Law Journal* vol. 34, 2019, pp. 754-755.

grandes plataformas y redes sociales para evadir el reconocimiento y el cumplimiento de los derechos laborales de estas personas, especialmente en lo que a la PRL se refiere.

Lo expuesto se plantea a meros efectos ilustrativos, porque el ejemplo que interesa en el presente estudio es el segundo: las dinámicas que plantean mantener como eje central la intervención humana pero que se ven absorbidas por el uso de la IA. El ejemplo que se plantea aquí es que un operador jurídico, si bien es el responsable de la diligencia debida en el ejercicio de su profesión, sea la abogacía o la judicatura, y, aunque solo pueda utilizar las herramientas automatizadas o de Inteligencia Artificial Generativa (IAG) como apoyo, caiga en un uso acrítico y negligente de las mismas.

En el ámbito que aquí ocupa, la pérdida de oportunidad real de apelación la sufrirá la parte del proceso social que no advierta un uso negligente de las herramientas de IA, que, por ejemplo, puede consistir en una fundamentación basada en alucinaciones generadas por un *chatbot* (errores jurisprudenciales, normativos o de cualquier índole jurídica). Sobre esto se ahondará más adelante.

El desajuste entre la apariencia del tipo de modelo de automatización de decisiones y la realidad del control humano sobre el mismo posibilita la ambigüedad sobre las dinámicas y obstaculiza su gobernanza. Por ese motivo, los autores citados proponen una conceptualización de las dinámicas como punto de partida del ámbito de la toma de decisiones automatizadas en general, conforme al tipo de desalineación: humanidad esqueuomórfica o automatización simulada, y, para cada caso, analizan las cuestiones normativas relacionadas a partir del teórico ideal y la discordancia con la realidad de la participación humana. Se plantea que cabe formular las siguientes cuestiones: ¿A qué tipo de dinámica nos enfrentamos? ¿Cuál es el estado final deseado? ¿Cómo debemos proceder dentro del contexto de esa dinámica?

En la problemática objeto de tratamiento nos enfrentamos a lo que la doctrina de referencia que pretende trasladarse denomina «segunda dinámica de la humanidad esqueumórfica», que se traduce en que lo ideal es que el sistema de toma de decisiones no debe estar ni parecer totalmente automatizado, pero lo está (o, se podría matizar a la luz de la casuística que se observa actualmente, está más automatizado de lo que debería, aunque no sea de forma total). En cierta medida, esta situación deriva de la complejidad de integrar estos sistemas en ámbitos que requieren cuestiones de gobernanza.

La aplicación de los dos principales problemas que la integración de los sistemas de toma de decisiones (en teoría, semi automatizados) se manifiesta de dos formas. La primera, el esfuerzo y la voluntad política que supone mantener o, dependiendo del caso, recuperar la integridad del factor humano a los actores que impulsaron la automatización puede ser contrario a sus intereses económicos. La segunda, aun en el caso de consenso sobre el carácter fundamental del factor humano, se suscitarán debates como en qué punto del proceso debería tener lugar la supervisión humana, quiénes y cómo debiera hacerse.

Resulta interesante pensar en cómo se manifiesta el esqueumorfismo que combina la automatización y el engaño en escenarios que provocan fricciones socialmente relevantes. En el concepto de la humanidad esqueumórfica la cuestión que preocupa es que se genere una tendencia a aceptar pasivamente formas de automatización desaconsejables. La desalineación entre la automatización real y la figurada puede impedir la ponderación de los intereses implicados y al mismo tiempo la búsqueda de soluciones requiere un tratamiento público y democrático.

Una delegación desaconsejable puede ocurrir en el caso de la jurimetría. Se trata de una herramienta que se utiliza para la predicción sobre el devenir procesal de los asuntos según sus características y la estadística del juzgador (con nombres y apellidos) [5]. Esta es una cuestión que ha de considerarse desde un punto de vista histórico evolutivo, porque si bien siempre han existido las quinielas *ad extra* de las inclinaciones de cada juzgador basadas en la práctica de los letrados, la creación de bases de datos que facilitan una calificación de la persona parece excesivo y puede influir en la sostenibilidad de la imparcialidad debida[6]. Paralelamente, una excesiva confianza en la predicción puede provocar errores estratégicos en la asistencia letrada. Por dicha razón, no resulta sorprendente que en Francia se haya prohibido publicar información estadística sobre las decisiones y el patrón jurisprudencial de las personas juezas, de forma que los datos de identidad de los juzgadores no pueden utilizarse para evaluar, analizar, comparar o predecir sus prácticas profesionales reales o presuntas (ello sin perjuicio de la valoración que quepa hacer de la proporcionalidad de la reprochabilidad de la conducta en los términos planteados: un delito que lleva aparejada una pena de prisión) [7].

Como conclusión contextual, estamos inmersos en un bucle de esqueumorfismo de los operadores jurídicos, posiblemente provocado por la exigencia del ritmo de trabajo tanto de los despachos como de los tribunales, en el que se proyecta una imagen de diligencia que realmente puede estar viciada por un uso acrítico de las herramientas de IA.

Como ejemplo de la casuística emergente, mediante el Auto del TSJ (Sala de lo Civil y Penal) de las Palmas de Gran Canaria, de 10 de febrero de 2026[8] se resolvió una pieza separada al asunto de origen tras detectar indicios de una conducta negligente en la interposición de un recurso. La mala *praxis* consistió en la inclusión sin revisión previa de numerosas citas, eminentemente jurisprudenciales y vertidas por un sistema de IA generalista, que no existían (lo que se conoce como «alucinaciones» de la IA). Incoada la pieza separada sobre mala fe procesal y tras la emisión del informe realizado por el Servicio de Documentación y Bibliotecas del TSJ de Canarias, que evidenció que de las 52 citas contenidas en el escrito solo 4 eran correctas y veraces, se acordó la imposición de una multa basando la decisión en las siguientes líneas argumentales.

En primer término, por vulneración de las máximas fijadas por el Libro Blanco sobre la IA y la Abogacía elaborado por el CGPJ junto al Ilustre Colegio de la Abogacía de Valencia, en tanto la IA constituye una herramienta para la eficiencia que no sustituye la labor del abogado y la diligencia

[5] Estadística sobre el número de asuntos resueltos y sobre el fallo por cada tipo de asunto (por ejemplo, en el orden social el porcentaje de casos resueltos a favor de la persona trabajadora y viceversa, etc.)

[6] En adhesión a lo expresado por el Magistrado Peralta Gutiérrez, «…la justicia predictiva no debe conducir al perfilado ideológico de jueces, ya que sus preferencias personales están protegidas por derechos fundamentales. El perfilado podría encasillar a los jueces, socavar su independencia y conducir a predicciones inexactas y discriminación, afectando su imparcialidad y privacidad». Peralta Gutierrez, A.: «Impacto comparado de la IA en la Administración de Justicia. La Guía Práctica de la Inteligencia Artificial», Nº 5, 2025, Aranzadi La Ley, p. 10.

[7] Véase el art. 33 de la *LOI n° 2019–222 du 23 mars 2019 de programmation 2018–2022 et de réforme pour la justice*. https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000038261631.

[8] En la misma línea casuística, léase la Nota Informativa Nº 90/2024 de la Oficina de Prensa del Gabinete del Presidente del Tribunal Constitucional. Disponible en: https://www.tribunalconstitucional.es/NotasDePrensaDocumentos/NP_2024_090/NOTA%20INFORMATIVA%20N%C2%BA%2090-2024.pdf

A este respecto, se considera que en ocasiones puede caerse en una desviación del punto de mira esencial por cuestiones que son periféricas a lo que resulta verdaderamente el fondo de los asuntos. Aquí quiere ejemplificarse con las razones que se esgrimieron a la hora de reprochar la conducta negligente de basar el escrito en alucinaciones jurisprudenciales. En ningún punto de la fundamentación de la Resolución ni de la Nota Informativa que posteriormente publicó el TC se hacía referencia a la salvaguarda de los intereses de la persona afectada por el asunto, sino únicamente al debido respeto hacia el Tribunal. No se pretende aquí menospreciar el debido respeto a Jueces y Tribunales, sino poner de manifiesto que la preocupación principal sobre este tipo de mala *praxis* debería ser la persona afectada y los derechos y garantías que les son inherentes.

y responsabilidad que le es inherente, que, en este caso, consiste en revisar si las referencias son correctas. Si bien no podía observarse al momento de los hechos por la entrada en vigor y por el ámbito subjetivo de aplicación, el TSJ de Canarias menciona, a título ilustrativo, el carácter preceptivo de la Instrucción 2/2026 del CGPJ sobre la utilización de sistemas de IA en el ejercicio de la actividad jurisdiccional por parte de la Carrera Judicial (en adelante, Instrucción).

Sentado el marco conceptual de los razonamientos jurídicos, el TSJ de Canarias califica la conducta del letrado en cuatro actuaciones reprochables. La primera, la dejadez en el empleo de una herramienta de IA generalista en lugar de emplear alguna de las diversas opciones existentes creadas *ad hoc* para la aplicación al ámbito jurídico (aunque podrá observarse mediante el apartado segundo del presente artículo que la diligencia debida no queda salvaguardada por el empleo de un sistema de IA jurídico). La segunda, la omisión de la debida revisión y verificación de los resultados que emanaron de la herramienta de IA, que el Tribunal califica de «quiebra del deber básico de supervisión humana». La tercera, la transgresión del deber de veracidad y de buena fe procesal a tenor de lo establecido por el art. 247.1[9] de la LEC[10], art. 61.1[11] del EGAE[12] en relación con el art. 10.2[13] del Código Deontológico y el art. 7.1. del CC. La cuarta, la falta de respeto y de consideración hacia el Tribunal *ex* art. 553 LOPJ en relación con el art. 247 de la LEC, debido a que, si bien se considera una acción negligente y no dolosa, de no haberse llevado a cabo la tarea de supervisión por parte del juzgador «hubiera dado lugar al cuestionamiento de la labor de este órgano de apelación, cuando no, si se hubieran dado por ciertas a la risión y descrédito absoluto de la Sala».

El cuarto razonamiento jurídico se dedica a la exposición de los dos precedentes conocidos hasta la fecha, un primero en el que la conducta negligente también se tradujo en el empleo de citas inexistentes atribuidas precisamente a la Sala del TC a la que iba dirigido el recurso de amparo y un segundo en el que pretendía aplicarse una norma del CP colombiano en lugar del español.

El quinto de los razonamientos se reserva a la graduación de la sanción y el sexto, si bien no responde al fondo del asunto, procede mencionarlo. El sexto de los razonamientos jurídicos se dedica a responder a una de las manifestaciones del letrado contenidas en el escrito por el que emitía sus disculpas. Concretamente, que pudo ser identificado por los compañeros debido al NIG, que, introducido en la base de datos permite la obtención de la resolución original que contiene los nombres de los profesionales que intervienen en el proceso (lo que no es objeto de anonimización, pues es una información pública). Lo que quiere subrayarse sobre dicho extremo es que, tal y como queda mencionado por el Tribunal, en el ámbito judicial se utilizan una serie de herramientas de IA suministradas por el CGPJ que componen el denominado KENDOJ, que sirve, entre otras acciones, para anonimizar las resoluciones de forma automatizada[14].

[9] «Los intervinientes en todo tipo de procesos deberán ajustarse en sus actuaciones a las reglas de la buena fe».

[10] «BOE» n. 7, de 08/01/2000.

[11] «Los profesionales de la Abogacía están obligados a respetar las normas deontológicas de la profesión contenidas en este Estatuto General, en el Código Deontológico de la Abogacía Española, en el Código Deontológico del Consejo de la Abogacía Europea y en cualesquiera otros que le resulten aplicables».

[12] «BOE» n. 71, de 24/03/2021.

[13] Si bien no se cita la letra, aquí se traen a colación las relacionadas con el asunto: «a. Actuar con buena fe, lealtad y respeto. b. Cumplir con los fines de la Administración de Justicia. c. Guardar respeto a todos cuantos intervienen en la Administración de Justicia exigiendo a la vez el mismo y recíproco comportamiento. e. Contribuir a la diligente tramitación de los asuntos que se le encomienden y de los procedimientos en los que se intervenga.»

[14] Según el portal de la administración de Justicia, también se utiliza el sistema ADIA y la única información publicada describe que «es una aplicación impulsada por IA, diseñada para optimizar el procesamiento de documentos. Ofrece las siguientes funcionalidades; anonimización: garantiza la privacidad y confidencialidad al eliminar datos sensibles, siendo ideal para procesos donde la protección de la información es esencial; resumen: genera una versión condensada del documento, destacando información clave para un acceso rápido y eficiente a los contenidos principales; lectura clara: transforma documentos jurídicos en versiones redactadas con un lenguaje claro y accesible, facilitando la

Sin perjuicio de esta sucinta referencia casuística, lo que todavía resulta más preocupante es la emergencia de conductas análogas por parte de los juzgadores[15]. Al respecto, es conveniente atender transversalmente a las advertencias de la Relatora Especial de la AGNU sobre la independencia de personas magistradas y abogadas, Margaret Satterthwaite (en adelante, «la Relatora»)[16].

Adviértase la gravedad del planteamiento de la Relatora sobre que, en última instancia, la erosión de la discrecionalidad judicial deriva en comportamientos mecanicistas[17] que socavan el pensamiento crítico y la eticidad de las prácticas, degradando las competencias propias de la investigación jurídica, de interpretación y redacción. Esto ha de observarse desde la perspectiva de que la judicatura tiene un poder epistémico que quedaría degradado, y si se relaciona con los actores poderosos que pueden cometer injerencias, puede manifestarse como otra forma de control. En efecto, la imparcialidad e independencia no pueden garantizarse si el juzgador se convierte en un elemento meramente ornamental, sino que requiere garantizar el acceso a un juez humano, de igual forma que a un abogado humano. Al mismo tiempo, la UNESCO[18], además de advertir sobre que el sesgo de automatización constituye una de las principales amenazas generadas por el empleo de sistemas de IA en la administración de justicia, sostiene que la adhesión acrítica de las decisiones basadas en IA puede provocar dinámicas de evasión de la motivación de las decisiones. Incluso, se plantea que puede llevar a considerarse que lo que merece justificación es la desviación respecto a los resultados que ofrece la herramienta. El sesgo de automatización dificulta el delimitar la línea que divide el campo entre lo que sirve de apoyo y lo que influye en las decisiones, por ejemplo, si de entrada un resumen no tiene en cuenta o minimiza ciertos elementos de forma que el juzgador termine obviándolos.

2. LA NECESIDAD DE UNA ALFABETIZACIÓN TÉCNICA

La justificación de este apartado radica en la observación de un tratamiento jurisprudencial y doctrinal meramente superficial del problema de las alucinaciones de *LLM*s, que van más allá de los casos más burdos debido a la complejidad que supone la integración de los modelos en el campo jurídico. Para ello se extraen conjuntamente las nociones facilitadas por dos estudios técnicos[19].

2.1. Diversidad taxonómica de los errores técnicos en el ámbito jurídico

Una de las primeras distinciones que ha de tenerse en cuenta es la diferencia entre los Modelos de IAG respecto a los modelos de IA consultiva, debido a que sus objetivos y procesos de funcionamiento, junto a los riesgos que entrañan, son distintos.

---

comprensión sin alterar el contenido legal esencial. Dirigida principalmente a Jueces, Fiscales y Letrados de la Administración de Justicia, ofreciendo soporte especializado para facilitar su labor diaria mediante el procesamiento inteligente de documentos». Disponible en: https://www.administraciondejusticia.gob.es/-/adia-anonimizado-de-documentos?inheritRedirect=true&redirect=.

[15] Véase la Sentencia T-323 de 2024 de la República de Colombia, Corte Constitucional-Sala Segunda de Revisión. En España se ha incoado un expediente disciplinario cuya resolución no está publicada por el CGPJ al momento de presentar el presente estudio, por lo que no procede su valoración, sino su mención a efectos de alertar sobre la existencia de este tipo de conducta negligente por parte del juzgador.

[16] Satterthwaite, M.: *Informe sobre la Independencia de los magistrados y abogados*. AGNU, 2025, pp. 1-23.

[17] A la hora de resumir escritos, buscar analogías entre asuntos, filtrar documentación jurídica de toda clase, predecir resultados, evaluar riesgos, obtener plantillas o borradores de sentencias, redactar providencias rutinarias, editar, corregir y revisar la ortografía y gramática de las resoluciones.

[18] UNESCO: *Kit De Herramientas global sobre IA y el Estado de Derecho para el poder judicial*. Francia: Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, pp. 99-100.

[19] Dantart, A.: «Inteligencia Artificial jurídica y el desafío de la veracidad: análisis de alucinaciones, optimización de RAG y principios para una integración responsable». *arXive-prints, arXiv:2509.09467*, *Cornell University*, 2025, pp. 1-118 y Magesh, V., *et al*: «Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools». *Journal of empirical legal studies*, 2025, vol. 22, n.º 2, pp. 216-242. https://arxiv.org/abs/2509.09467.

El objetivo de la IAG en el contexto de los *LLMs* no es aportar veracidad frente al *prompt* (solicitud/pregunta) que se introduce, sino mantener un hilo conversacional y una coherencia probabilística. Estos modelos nutridos de una ingente cantidad de datos, predicen las palabras más probables en una secuencia de patrones estadísticos que se desprenden de su entrenamiento (no razonan). El problema radica en que las alucinaciones, a diferencia de lo que pueda pensarse por los *legos* en programación, no son un mero error a corregir, sino que precisamente forman parte de su naturaleza, que no deja margen a la incertidumbre o a una respuesta que indique desconocimiento respecto a la pregunta[20]. Es por ello que este tipo de herramientas son útiles para tareas creativas, pero no para aquellas que requieren una precisión factual, como es el ámbito jurídico. A esto hay que añadir que la característica de caja negra impide, incluso a los profesionales especializados, acceder a la debida trazabilidad de los orígenes del resultado. A su vez, cuando es un modelo de propósito general, puede retroalimentarse con contenido de baja calidad que, como bien es sabido, amplifica y perpetúa sesgos, lo que da lugar a resultados discriminatorios que mantienen injusticias sistémicas.

Por lo tanto, tratar de utilizar un *chatbot* de IAG como apoyo jurídico sabiendo que la alucinación es una característica intrínseca a su funcionamiento, no parece razonable. De la serie de ejemplos que se exponen en el informe técnico de referencia, cabría detallar que, si bien puede resultar útil para la creación de borradores en otro tipo de profesiones, no lo será por el momento en el ámbito jurídico.

Por su parte, la IA consultiva no se centra en generar contenido, sino en recuperarlo, estructurarlo y presentarlo una vez verificado. Básicamente, su arquitectura se construye sobre el sistema de Generación Aumentada por Recuperación (RAG). Este se nutre de una base de datos que puede y, en el contexto jurídico, debe, ser continuamente actualizada. Para su tratamiento, el sistema RAG divide la información en unidades manejables (*chunks*) y se generan *embeddings* vectoriales (representaciones del significado semántico de los *chunks*) almacenándose en una base de datos vectorial. Por otro lado, las consultas también se dividen en *embeddings* vectoriales, para que cuando el usuario introduzca su consulta, se busquen las similitudes semánticas. Es por ello que la fase de recuperación de la información no puede limitarse, al menos en el ámbito jurídico, a la búsqueda de términos semánticamente similares o documentos, sino a traer a colación lo que sea jurídicamente y contextualmente pertinente, en su justa medida, pero con exhaustividad y concisión (de ahí la necesidad de una comprensión profunda jurídico-técnica). Una de las principales fortalezas de su diseño es que facilita una trazabilidad de las citas al corpus documental, que es en lo que, por el momento, se traduce la diligencia debida del operador jurídico que lo utiliza. Otra de las claves prometedoras de este sistema es poder especializar al máximo tales significados al ámbito jurídico con la complejidad que ello conlleva. El técnico del informe que se ha tomado como referencia lo ejemplifica con el problema que supone que la profundidad semántica del sistema no abarque cuestiones esenciales para una buena interpretación jurídica, como pueda ser la jerarquía normativa, la vigencia normativa, etc. Concretamente plantea que, si es entrenado bajo los parámetros de la pirámide de Kelsen, podrán depurarse errores del tipo jerárquico.

En esencia, mientras los modelos de IAG trabajan sobre un corpus cerrado que depende del entrenamiento de retroalimentación, que queda estanco en cuanto a las fuentes, los sistemas RAG funcionan como un libro abierto en cuanto al control de las fuentes, ya que se nutren de una base de datos trazable y actualizable.

[20] «…los modelos de lenguaje están optimizados para ser buenos en la realización de pruebas, y adivinar cuando no se está seguro mejora el rendimiento en las pruebas. Esta "epidemia" de penalización de las respuestas inciertas solo puede abordarse mediante una mitigación sociotécnica: modificar la puntuación de los puntos de referencia existentes que están desalineados, pero dominan las clasificaciones, en lugar de introducir evaluaciones adicionales de alucinaciones. Este cambio podría orientar el sector hacia sistemas de IA más fiables». Kalai, A.T., *et al*.: «Why language models hallucinate». *arXiv preprint Cornell University, arXiv:2509.04664*, 2025.

Sentada esta distinción, resulta imprescindible conocer que las alucinaciones legales ya tienen su propia taxonomía, que va mucho más allá de una calificación como resultado correcto o incorrecto. Siguiendo la clasificación de Dantart y Magesh, básicamente, los tipos de alucinaciones jurídicas se encuadran conforme a cuatro categorías principales: factuales/extrínsecas («*Misstatement* de Ley o precedente»[21], «fabricación de autoridad»[22] y «error de aplicación jurisdiccional/temporal»[23]) basadas en fuentes («fundamentación errónea/*misgrounding*»[24] y «falta de fundamentación/*ungrounding*»[25]), de inferencia y razonamiento («argumentación ilógica o inválida»[26] y «*miscaracterización* de argumentos/partes»[27]), e intrínsecas («inconsistencias en el *prompt* o corpus de entrenamiento»). Por otro lado, con la «supresión de citas problemáticas» se hace referencia a los errores que consisten en afirmaciones del sistema basadas en resoluciones derogadas que se dan por buenas por no incluir la cita.

A mayor abundamiento, no solo es fundamental conocer la tipología de las alucinaciones, sino los distintos tipos de sesgos específicos que emanan de su aplicación a lo jurídico y que vician los resultados que arrojan[28]. Por ejemplo, el «sesgo de primacía» por el que se centra desproporcionadamente en la información que se recoge al inicio de un documento o el «sesgo de centralismo», por el que se fija en exceso en los partidos judiciales con mayor volumen de asuntos ignorando las especificidades de otros territorios. Resulta igualmente necesario advertir los «sesgos contrafácticos», por los que el propio modelo acepta acríticamente premisas falsas contenidas en las demandas del usuario (*prompts*).

2.2. Sobre los sistemas RAG

Consecuentemente, si se observa el funcionamiento de los sistemas RAG junto a la tipología de errores y sesgos en los que caen los modelos, resulta sencillo pensar en el riesgo que suponen ejemplos como los errores de interpretación que puede provocar el dividir una sentencia en *chunks* y establecer correlaciones semánticas, obviando que en un fundamento puede parecer que el juzgador se inclina hacia determinada posición, pero en los siguientes defiende lo contrario. En consecuencia, se altera el significado de las resoluciones.

Magesh[29] advierte que los sistemas RAG se componen de múltiples componentes discretos que pueden ser cajas negras y que la opacidad de los proveedores impide conocerlos completamente o llegar a descifrar por qué alucinan sobre una cuestión y no sobre otra, así como que no se puede

[21] Descripción incorrecta del contenido normativo o jurisprudencial existente, desde sutiles tergiversaciones hasta contradicciones notorias.

[22] Invención completa de casuística, normativa e incluso operadores jurídicos o personas investigadoras.

[23] Aplicación incorrecta de los principios correspondientes a otra jurisdicción, citación de normativa derogada o confusión del carácter de la aplicación temporal o espacial de la misma.

[24] Citación de referencias normativas o casuísticas reales tergiversando su contenido o aplicándolo incorrectamente.

[25] Afirmaciones factuales sin fundamentación, que omiten la referencia a las fuentes de donde se ha recuperado la información.

[26] Razonamientos que quiebran principios lógicos o sin base jurídica, que superficialmente pueden resultar persuasivos.

[27] Confusión de posiciones procesales, las partes y sus respectivas argumentaciones.

[28] En sentido contrario, la Relatora pone de manifiesto que los sistemas de IA también se pueden programar para detectar los sesgos de los juzgadores, pudiendo además facilitar una perspectiva multidisciplinar que enriquezca sus conocimientos (lo ejemplifica con un proyecto que aporta perspectivas de neurociencia y psicología infantil para los casos de violencia y propone decisiones que no caigan en las propias de estructuras patriarcales). De igual forma, los sistemas pueden señalar el uso de lenguaje despectivo. Para lograrlo, el diseño debe ser exhaustivo y debe ser objeto de revisión continua.

[29] Estudio de campo de referencia que desmiente las afirmaciones de algunos proveedores de los sistemas de IA que aseguran mitigar o eliminar por completo el riesgo de las alucinaciones. Magesh, V., *et al*: «Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools». *Journal of empirical legal studies*, 2025, vol. 22, n.º 2, pp. 216-242.

establecer un único tipo de error para cada alucinación. Como se deduce de lo expuesto, los sistemas RAG son prometedores, pero siguen alucinando y requieren el mismo esfuerzo revisor por parte de los operadores jurídicos. En este sentido, Magesh nos proporciona evidencias suficientes sobre que los resúmenes automatizados de las sentencias «están plagados de incoherencias y alucinaciones»[30].

También tenemos a nuestra disposición otros estudios de campo que evidencian que los modelos RAG y los *LLM* no posibilitan una generación de resúmenes totalmente automatizada y fiable en un ámbito tan complejo como el Derecho y que dependen de la intervención y supervisión humana especializada (no solo para el diseño e implantación de los modelos, sino para la compleja detección de errores)[31]. Esto también suscita cuestionarse si surgirá un trabajo atípico con el fin de cubrir dicha necesidad.

El entrenamiento de las herramientas de IA en un corpus cerrado de legislación nacional (como se plantea con KENDOJ) puede mejorar su desempeño, pero además de requerir bases de datos jurídicas locales actualizadas, necesita entrenar los mencionados *embeddings* porque una simple relación semántica no sirve para dar respuestas jurídicas correctas.

Por todo lo expuesto, resulta llamativo que algunas voces insistan en la confianza sobre el empleo de la base de datos CENDOJ para transmitir la tranquilidad de que la información que manejan los sistemas KENDOJ se puede dar por buena, no solo por los problemas estructurales predichos, sino porque, como advierte la Relatora, el hecho de que los modelos se nutran de la jurisprudencia anterior entraña el riesgo de socavar la discrecionalidad judicial hacia el sesgo de automatización. Que un modelo se desarrolle internamente no exime de las discriminaciones y sesgos estructurales, en muchas ocasiones, por la falta de una suficiente mirada interdisciplinar.

Además, contamos con un estudio que audazmente cuestiona si las sentencias son un dato idóneo para alimentar los sistemas de IA, teniendo en cuenta que pueden contener errores, estar sesgadas, haber provocado efectos discriminatorios, carecer de motivación suficiente, etc. Este alcanza la misma conclusión que se obtiene del presente estudio, en tanto no se puede considerar la sentencia como dato de calidad *per se*, y que, por lo tanto, ese automatismo no puede ocurrir[32]. A esto cabría añadir la necesidad de evitar un estancamiento jurisprudencial y el análisis de dónde poner la línea temporal de los datos históricos a tener en cuenta, para que no se incorporen líneas superadas cuya aplicación no tiene ningún sentido en la actualidad.

En definitiva, siguiendo a Dantart y a diferencia de lo que intentan trasladar algunos proveedores, los sistemas RAG no son la panacea y requieren de una ingeniería de datos y refinamiento continuos. Si bien la arquitectura de los sistemas RAG reduce significativamente la probabilidad de caer en alucinaciones notorias, sí puede caer en alucinaciones sutiles o, lo que los técnicos denominan «insidiosas» en la complejidad del ámbito jurídico, porque provocan una falsa sensación de fiabilidad y su detección se dificulta si se descuida la debida verificación de cada fuente citada. A este respecto, queda evidenciado que gracias a estudios con el enfoque multidisciplinar técnico-jurídico se puede conocer mejor la taxonomía de estas alucinaciones jurídicas, que no pueden reducirse a una calificación de correcto/incorrecto.

[30] «LexisNexis (Lexis+AI) y Thomson Reuters (Westlaw AI-Assisted Research y Ask Practical Law AI) alucinan entre el 17 % y el 33 % de las veces» y «Ask Practical Law AI de Thomson Reuters proporciona respuestas incompletas (negativas o respuestas sin fundamento… en más del 60% de nuestras consultas, la tasa más alta entre los sistemas que hemos probado».

[31] Deroy, A.; Ghosh, K., Ghosh, S.: «How ready are pre-trained abstractive models and LLMs for legal case judgement summarization?» *arXiv preprint arXiv:2306.01248* Cornell University, 2023.

[32] Juliá Pijoan, M. (2024): *La computarización del derecho: a partir del proceso y de los procedimientos judiciales*, Madrid: Dykinson, pp. 97-119.

2.3. Problemas subyacentes que amenazan la integridad de los procesos judiciales en general

El informe de la Relatora especial tiene un carácter holístico, que también advierte sobre las consecuencias que puede tener el desarrollo monopolístico de estas herramientas por parte del sector privado. Esto puede afectar a la independencia judicial y privar a la judicatura de la información necesaria para evaluar los modelos, su desarrollo y entrenamiento, junto a los riesgos de error y sesgo que pueden albergar mientras están protegidos por patentes o secretos comerciales. Con esto pretende transmitirse que no solo es necesario conocer el funcionamiento, la arquitectura y los riesgos de los sistemas, sino las cuestiones de economía política subyacentes. Tomando el testigo del planteamiento de la Relatora sobre el panorama desastroso que puede suponer que actores poderosos instrumentalicen estas herramientas para la consecución de sus objetivos, a juicio de la autoría del presente artículo, resulta necesaria una mirada desde la Economía Política de la IA, que analice críticamente las relaciones de poder y la dirección de los sistemas, que pueden socavar la independencia del poder judicial[33].

La Relatora también advirtió sobre ciertos escenarios que vulneran la independencia, ante medios que, mal empleados, podrían utilizarse para presionar a los jueces a seguir las recomendaciones generadas mediante los modelos que proporcionan decisiones asistidas o automatizadas (por ejemplo, mediante la referida imposición de indicadores del desempeño u objetivo de «eficiencia»). En consecuencia, otro de los puntos esenciales a abordar sobre la integridad de los procesos, además de la desconfianza pública en la justicia que puede generar el empleo de ultrafalsificaciones[34], tal y como lo contextualiza la Relatora, es que las personas juezas y magistradas son objeto de ataques por parte de distintos actores poderosos y las herramientas que se tratan aquí no quedan al margen del tecnoautoritarismo.

En esta línea, otra de las consideraciones fundamentales de la Relatora, que puede pasar desapercibida si se aborda la cuestión como si fuese una realidad estanca o sin elementos causales de carácter político, es el de analizar el tipo de poder epistémico subyacente. Esta forma de operar está derivando en un poder epistemológico que se traduce en la capacidad de definir qué cuenta como conocimiento en una disciplina. Un ejemplo que puede ser asumido de forma acrítica es que la Instrucción sobre el empleo de la IA haga referencia a la «doctrina relevante» (ordinal 6º.3.a): ¿qué se entiende por doctrina relevante? ¿la mayoritaria? ¿entrenamos los modelos de IA para que cada vez se concentren más sobre esas inclinaciones? ¿dónde queda el margen para considerar otras líneas?

La Relatora trae a colación el principio de igualdad de medios procesales y la consiguiente capacidad de impugnar las pruebas presentadas de contrario, advirtiendo que la vulneración de dicho principio puede suponer que, por ejemplo, la Administración emplee sistemas con códigos fuente que no puedan ser objeto de impugnación por quedar protegidos por los derechos asociados a la tecnología patentada o el secreto comercial. Veremos que esa protección que opaca[35] la información que puede ser objeto de impugnación por las partes procesales ya no puede sostenerse. Y como claramente manifiesta la Relatora, «no hay forma de impugnar la idoneidad de las tecnologías si no constan en el expediente judicial». Tal y como advierte la Relatora, los actos procesales pueden afectar a los

[33] Sirva de antecedente el supuesto expuesto por la Relatora sobre que en algunos países que utilizan modelos de IA para supervisar y evaluar las decisiones de los juzgadores llegando a incentivar la inclinación conforme a los resultados que estos arrojen.

[34] Uno de los principales problemas que advierte la Relatora y que es esencial de cara a la reflexión que pretende suscitarse mediante el presente trabajo es la relación de problemas asociados a la impugnación de pruebas generadas o manipuladas mediante IA.

[35] La opacidad ha de interpretarse desde una mirada tridimensional: debida al secreto corporativo, al analfabetismo técnico y a las características inherentes a los algoritmos de aprendizaje automático. Juliá Pijoan, M. (2024): *La computarización del derecho: a partir del proceso y de los procedimientos judiciales*, Madrid: Dykinson, p. 81.

DD.FF. tanto como las decisiones sustantivas y lo que resulta más peligroso en la temática objeto de tratamiento es la carencia de vías de recurso significativas (o podría decirse, eficaces). Además, el coste que puede suponer adquirir herramientas especializadas puede exacerbar la desigualdad de medios entre las partes creando un ecosistema jurídico estratificado. Las herramientas de IA predictiva previamente referidas sobre las que comienzan a apoyarse las estrategias procesales son un ejemplo de ello.

3. PRINCIPALES PROBLEMAS DETECTADOS EN LA INSTRUCCIÓN 2/2026 DEL CGPJ, SOBRE LA UTILIZACIÓN DE SISTEMAS DE IA EN EL EJERCICIO DE LA ACTIVIDAD JURISDICCIONAL[36]

La Instrucción tiene por objeto marcar los métodos y principios que han de respetarse en el uso de las herramientas de IA por parte de los juzgadores, observando en todo caso lo establecido por la LOPJ y resto de normativa aplicable[37]. La citada Instrucción es de obligado cumplimiento para los jueces, juezas, magistrados y magistradas en lo relativo a los sistemas de IA[38] (incluidas las herramientas de IAG[39]) que puedan utilizar para el ejercicio de la actividad jurisdiccional *ex* art. 2 de dicho texto.

Sin perjuicio de la interpretación teleológica de los límites del apoyo en las herramientas de IA para la toma de decisiones, la Instrucción, salvo error interpretativo de quien suscribe estas líneas, contiene algunos apartados normativos que resultan contradictorios.

De los cuatro bloques de usos que quedan permitidos mediante el ordinal sexto, a efectos de lo que aquí interesa procede centrar el foco especialmente en la letra c) de su apartado tercero. La letra c) establece que se podrán utilizar sistemas de IA (facilitados por la Admón.) para la «elaboración de esquemas, resúmenes o borradores de trabajo interno, siempre que no tengan carácter decisorio ni sustituyan la redacción personal de resoluciones judiciales». El apartado cuarto del mismo ordinal, establece que «la utilización de sistemas de IA para la elaboración de borradores de resoluciones judiciales únicamente podrá realizarse en los términos y con los límites establecidos en el ordinal siguiente». El elemento contradictorio viene precisamente de la mano del siguiente ordinal, el séptimo, apartado cuarto, cuyo tenor literal es el siguiente: «A los efectos previstos en los apartados anteriores, los sistemas de IA que se utilicen en el ejercicio de la función jurisdiccional deberán asegurar que el borrador documental sólo se genere a voluntad del juez, jueza, magistrado o magistrada y pueda ser libre y enteramente modificado por estos antes de su validación como resolución judicial o procesal».

Esta redacción abre la posibilidad de que el borrador creado con IAG pueda validarse como resolución judicial o procesal, ya que atribuye el hecho de modificar el resultado del sistema de IA

[36] «BOE» n. 27, de 30 de enero de 2026, páginas 14480 a 14487 (8 págs.)

[37] Junto a lo que constituyó uno de los pasos originarios sobre la materia: la Política de uso de la IA en la Administración de Justicia adoptada por el Comité Técnico Estatal de la Administración Judicial Electrónica (regulado en el Real Decreto 396/13, de 7 de junio, con la finalidad de favorecer la compatibilidad y asegurar la interoperabilidad de los sistemas y aplicaciones empleados en la Administración de Justicia, así como la cooperación entre las distintas Administraciones). Consúltese en: https://www.administraciondejusticia.gob.es/documents/7557301/7558184/CTEAJE-NOR-Politica+de+uso+de+la+IA+en+la+AJ+v1.0.pdf/ddc0eda1-950b-e926-b367-be511b16f2f9?t=1733925140122.

[38] «Todo sistema basado en máquinas que, de conformidad con el Reglamento de Inteligencia Artificial, esté diseñado para funcionar con distintos niveles de autonomía y que puede mostrar capacidad de adaptación tras el despliegue, y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales» (art. 3.a) Instrucción).

[39] «Aquellos sistemas de IA capaces de generar de forma automatizada contenidos, incluidos textos, imágenes, sonidos, código u otros materiales, a partir de instrucciones formuladas por el usuario» (art. 3 b) Instrucción).

a una opción que tiene el juzgador, no una obligación. Esto es, se prevé que el juzgador decida si modifica o no el borrador vertido por el sistema, sin obligar a construir enteramente la resolución. De esta forma, su uso no ha quedado delimitado en un mero apoyo que suponga un ahorro de tiempo para cuestiones formales, estructurales o repetitivas.

Otro punto que llama la atención es el de la rotunda afirmación del apartado tercero del ordinal séptimo, que afirma que los borradores generados por los sistemas de IA «no tendrán en ningún caso la consideración de decisiones automatizadas». Pues bien, si el juzgador, viciado por el sesgo de automatización, da por válido el borrador emitido y lo valida como resolución, parece evidente que será una decisión automatizada porque el razonamiento no lo ha llevado a cabo de forma autónoma, sino que es fruto del código del modelo. Por lo tanto, este apartado puede suponer un problema de cara a intentar depurar responsabilidades por una mala *praxis* y depurar el propio proceso, ya que conforme a la arquitectura normativa que se propone, se está abriendo la vía de la generación de decisiones automatizadas (a lo que hay que añadir, que no puntualiza si se refiere a decisiones total o parcialmente automatizadas).

El régimen sancionador al que se refiere el ordinal undécimo de la Instrucción[40], a juicio de la autoría del presente estudio evidencia una inseguridad jurídica desde las dos perspectivas que quedan involucradas: las partes procesales que detecten la mala *praxis* a efectos de sus alegaciones y súplicas, junto a los propios juzgadores que no disponen del desarrollo que la especificidad de la materia requiere. La lentitud y prosperidad de la aplicación del referido precepto, junto a los puntos susceptibles de desarrollo, provocan el panorama del que tanto se abusa: reactivo en lugar de preventivo.

Otra de las cuestiones que no parece conveniente es la dispuesta por el ordinal decimotercero de la Instrucción, en tanto prevé la formación y capacitación de personas juezas y magistradas de forma dispositiva, en lugar de preceptiva, lo que podría generar una dejadez y consecuente mala utilización de los mismos[41]. Esto se desvía del mandato del art. 26.2[42] del RIA en relación con su cdo. 73[43], que reserva la tarea de supervisión a personas formadas sobre la materia.

[40] «El incumplimiento de los criterios, pautas de uso y principios establecidos en la presente instrucción por los jueces, juezas, magistrados y magistradas podrá dar lugar a las responsabilidades que procedan, conforme a lo previsto en la Ley Orgánica del Poder Judicial, sin perjuicio de las medidas preventivas o correctoras que puedan adoptarse en el ámbito organizativo o de gobierno del Poder Judicial». La referencia a la falta de motivación en el régimen sancionador de la LOPJ es la del art. 417.15, que califica como falta muy grave «la absoluta y manifiesta falta de motivación de las resoluciones judiciales que la precisen, siempre que dicha falta haya sido apreciada en resolución judicial firme. Si la resolución inmotivada no fuese recurrible, será requisito para proceder la denuncia de quien fue parte en el procedimiento».

[41] En palabras de la Relatora sobre el sesgo de automatización: «se puede confiar acríticamente en los resultados de la IA si los jueces no reciben formación y orientación adecuadas sobre los problemas y las limitaciones de estas tecnologías, lo que socava la capacidad del enfoque basado en mantener a un ser humano para supervisar las decisiones de la IA».

[42] «Los responsables del despliegue encomendarán la supervisión humana a personas físicas que tengan la competencia, la formación y la autoridad necesarias».

[43] «Los sistemas de IA de alto riesgo deben diseñarse y desarrollarse de tal modo que las personas físicas puedan supervisar su funcionamiento, así como asegurarse de que se usan según lo previsto y de que sus repercusiones se abordan a lo largo del ciclo de vida del sistema. A tal fin, el proveedor del sistema debe definir las medidas adecuadas de supervisión humana antes de su introducción en el mercado o puesta en servicio. Cuando proceda, dichas medidas deben garantizar, en concreto, que el sistema esté sujeto a limitaciones operativas incorporadas en el propio sistema que este no pueda desactivar, que responda al operador humano y que las personas físicas a quienes se haya encomendado la supervisión humana posean las competencias, la formación y la autoridad necesarias para desempeñar esa función. También es esencial, según proceda, garantizar que los sistemas de IA de alto riesgo incluyan mecanismos destinados a orientar e informar a las personas físicas a las que se haya asignado la supervisión humana para que tomen decisiones con conocimiento de causa acerca de si intervenir, cuándo hacerlo y de qué manera, a fin de evitar consecuencias negativas o riesgos, o de detener el sistema si no funciona según lo previsto».

Con arreglo al art. 6.2 del RIA en remisión al anexo III, apartado 8º.a) del mismo cuerpo normativo, «los sistemas de IA destinados a ser utilizados por una autoridad judicial, o en su nombre, para ayudar a una autoridad judicial en la investigación e interpretación de hechos y de la ley, así como en la garantía del cumplimiento del Derecho a un conjunto concreto de hechos, o a ser utilizados de forma similar en una resolución alternativa de litigios»[44] se califican como sistemas de IA de alto riesgo. El estudio de la Universidad de Oxford que se cita[45], valora cómo puede materializarse la obligación de supervisar el empleo de sistemas de IA de alto riesgo por parte de los juzgadores. Este concluye que no se puede confiar en que una persona jueza promedio utilice y supervise al mismo tiempo sistemas de IA de alto riesgo para el apoyo en la toma de decisiones respetando lo establecido por el art. 14 del RIA[46]. Los sistemas que se utilicen en el ámbito judicial necesitan contar con garantías estructurales de forma previa a su despliegue para preservar la independencia, autonomía e imparcialidad de los juzgadores. En definitiva, dicho estudio alcanza la conclusión de que la tarea de supervisar los sistemas de IA de alto riesgo para el apoyo en la toma de decisiones judiciales no resulta asequible, cuando además, la clave está en abordar las cuestiones sociotécnicas que subyacen a los sistemas de forma interdisciplinar. En la misma línea, la Relatora sostiene que la incapacidad del juzgador de conocer los procesos que sigue el modelo de la herramienta de IA que utiliza para la toma de decisiones dificulta su motivación y apelación.

4. TRADUCCIÓN DE LA INSEGURIDAD JURÍDICA QUE PROVOCA ESTA REALIDAD EN EL PROCESO SOCIAL

En aras de evitar dilaciones indebidas y resolver los asuntos que atañen a la esfera del trabajo y la seguridad social de las personas trabajadoras, el proceso social se rige conforme a unos profundamente asumidos principios de oralidad, concentración y celeridad consagrados en el art. 74.1 de la LRJS, que no son absolutos, sino que «orientarán la interpretación y aplicación de las normas procesales propias de las modalidades procesales» a tenor de lo establecido por el segundo apartado del mismo precepto.

No obstante, por parte de la autoría de la presente investigación se considera que la irrupción de la IA sin unas técnicas de detección y depuración de la mala *praxis*, actualmente genera una inseguridad jurídica grave. Si bien las especificidades del proceso social tienen como fin último no generar dilaciones en una jurisdicción que ha de velar por asuntos cuya resolución requiere especial celeridad, el ritmo de la proposición, admisión, impugnación, práctica y valoración de la prueba junto a la unidad de acto, también puede provocar inseguridad jurídica. Es por ello que a juicio de la

[44] «dado que pueden tener efectos potencialmente importantes para la democracia, el Estado de Derecho, las libertades individuales y el derecho a la tutela judicial efectiva y a un juez imparcial… La utilización de herramientas de IA puede apoyar el poder de decisión de los jueces o la independencia judicial, pero no debe substituirlas: la toma de decisiones finales debe seguir siendo una actividad humana» (cdo. 61 RIA). En la misma línea, cdos. 48 y 58 del RIA sobre la magnitud consecuencias adversas sistema de IA para los DD.FF. protegidos por la CDFUE.

[45] Banks, I.: «Judges-in-the-loop? Judicial involvement in human oversight of high-risk decision support systems under the EU AI Act». *International Journal of Law and Information Technology. Oxford University Press*, 2026, 34 pp. 1-23.

[46] «a) entender adecuadamente las capacidades y limitaciones pertinentes del sistema de IA de alto riesgo y poder vigilar debidamente su funcionamiento, por ejemplo, con vistas a detectar y resolver anomalías, problemas de funcionamiento y comportamientos inesperados; b) ser conscientes de la posible tendencia a confiar automáticamente o en exceso en los resultados de salida generados por un sistema de IA de alto riesgo ("sesgo de automatización"), en particular con aquellos sistemas que se utilizan para aportar información o recomendaciones con el fin de que personas físicas adopten una decisión; c) interpretar correctamente los resultados de salida del sistema de IA de alto riesgo, teniendo en cuenta, por ejemplo, los métodos y herramientas de interpretación disponibles; d) decidir, en cualquier situación concreta, no utilizar el sistema de IA de alto riesgo o descartar, invalidar o revertir los resultados de salida que este genere; e) intervenir en el funcionamiento del sistema de IA de alto riesgo o interrumpir el sistema pulsando un botón de parada o mediante un procedimiento similar que permita que el sistema se detenga de forma segura.»

autoría del presente estudio no ha resultado sorpresiva la aprobación de la Ley Orgánica 1/2025, de 2 de enero, de medidas en materia de eficiencia del Servicio Público de Justicia, por la que se modifica, entre otros preceptos, el art. 82 de la LRJS y a tenor del cual la prueba documental o pericial de la que las partes intenten valerse habrá de proponerse con diez días de antelación al acto del juicio, salvo las tres excepciones lógicas que imposibilitan anticiparla. Cabe subrayar que esta modificación supone una apertura del ámbito objetivo de la anticipación de la prueba que con anterioridad se preveía para supuestos en los que el volumen de la prueba o su complejidad lo aconsejase. De esta forma, la excepcionalidad ahora es no aportar la prueba anticipada. También parece positiva la aportación en formato electrónico, debido a que, aunque suponga algo farragoso para las partes, puede facilitar el control del empleo de la IA en las pruebas que se aporten.

Nores Torres[47] estima este cambio como una modificación muy positiva en aras de mejorar la preparación del juicio, sobre todo, en el contexto digitalizado actual. De hecho, plantea directamente la procedencia de adoptar una audiencia previa y/o una contestación escrita en la línea de lo que sucede en otros sistemas[48]. Como es conocido, el posicionamiento contrario se traduce en considerar que se está civilizando el proceso social y que estas cuestiones vulneran los principios vertebradores del proceso.

En este punto se quiere visibilizar la traducción de las amenazas previamente expuestas en el proceso social. En la vista oral, sin perjuicio de la preceptividad actual de proponer la prueba con diez días de antelación, la admisión, impugnación y resolución se dan en el acto. Si concurre un caso de mala *praxis* en el uso de herramientas de IA, ¿se puede garantizar su detección y la depuración del proceso, junto al ejercicio de los correspondientes derechos de la parte afectada, si se mantienen a toda costa los principios clásicos? Teniendo en cuenta todo lo expuesto hasta el momento, parece evidente que no.

La doctrina de Nores Torres abre el debate general (sin especificar el supuesto que aquí se trae a colación), expresando que puede suponer un problema de quiebra de los principios, aunque no se trate de principios absolutos sino de una aspiración del sistema procesal. Lo que aquí se plantea es la referida reflexión sobre si prima la celeridad del proceso o la prevención de las consecuencias que puede tener sobre el devenir del asunto una mala *praxis* en el uso de la IA. Y se considera que, en cualquier caso, si ha de ponderarse la priorización de la celeridad o la seguridad jurídica, puede considerarse que prima, ante las amenazas de un empleo negligente de la IA, que se siga un proceso con las debidas garantías, que, no tiene por qué ser de una duración mayor. Precisamente, se plantea como una cuestión preventiva.

4.1. ¿Recurso de reposición? ¿Constituye una verdadera garantía teniendo en cuenta la celeridad del proceso?

Vistas las formas mediante las que puede quedar viciado el proceso, ¿los cauces que tiene la parte afectada por un uso negligente de la IA constituyen una garantía real? Principalmente, se pueden dar dos escenarios: el primero, que la parte afectada por las alucinaciones o falsedades de la contraparte no sea consciente de las mismas, porque revistan una apariencia de veracidad, y porque, como consecuencia de la estructura del proceso, resulte imposible verificarlo suficientemente (en

[47] Nores Torres, L.: «El orden social y el proceso laboral tras la LO 1/2025, de 2 de enero». *Trabajo y Empresa. Revista de Derecho del Trabajo* Vol. 4, n. 1. 2025. pp. 50-52.

[48] En la misma línea, una cuestión no introducida y que podía ser de interés según Tascón López: «como excepción al principio de oralidad, la contestación a la demanda pueda ser escrita (arts. 141 y 143 LRJS)». Tascón López, R.: «Algunas reflexiones sobre la situación del proceso social tras la Ley Orgánica 1/2025, 2 de enero, de medidas en materia de eficiencia del servicio público de justicia». *Revista Crítica de Relaciones de Trabajo, Laborum*. n 15º, vol. 2º, 2025, pp 29–52.

unos minutos y sin poder utilizar ninguna herramienta de apoyo, añadido a la tensión que supone el acto). Consecuentemente, en ese escenario los vicios quedarían insertos en la valoración de la prueba y se arrastrarían hasta la finalización del proceso, pudiendo influir en la decisión de la persona juzgadora. El segundo supuesto, en caso de que la parte afectada tuviese indicios razonables de la existencia de vicios, interpondría el recurso de reposición de la mejor manera posible teniendo en cuenta la referida falta de medios técnicos oportunos y de tiempo, se resolvería al momento por una persona que tampoco dispone de los recursos que requiere dicha situación, lo cual constituye una inseguridad jurídica muy grave.

Respecto al arrastre de pruebas falsas manipuladas por IA, las vías procesales parecen a primera vista «sencillas»[49]: impugnar en el acto de la vista activando la prejudicialidad penal (salvo que no se trate de una prueba fundamental) que resolverá sobre la falsedad documental o estafa procesal (si se descubriera tras la firmeza de la sentencia primero habría de acreditarse mediante el proceso penal y después interponer un recurso extraordinario de revisión). No obstante, nada de esto se activa si no se tiene la ocasión de detectarlo para impugnarlo en el acto del juicio (cuya fundamentación, improvisada y sin el sustento de una pericial, quedaría bastante pobre) y es sobradamente conocido que las ultrafalsificaciones sofisticadas son difíciles de detectar incluso para los técnicos[50]. ¿Todo esto es garantista?

4.2. ¿Qué encaje y valoración tendría el sesgo de automatización en el juzgador si se detectase a tiempo para ulteriores recursos?

En primer término, quedamos obligados a traer el mismo razonamiento: las ulteriores vías de impugnación también pueden quedar anuladas por el supuesto en el que un desconocimiento sobre los vicios que arrastra el proceso no genere las correspondientes alegaciones para su depuración o, en caso de ser detectadas, que sea demasiado complejo encuadrar la vulneración de la tutela judicial efectiva. En cualquier caso, si se detectase la mala *praxis* vulneradora de DD.FF. y generadora de indefensión, ¿prosperarían las vías que intentarían activarse en caso de detectar el vicio? ¿en qué motivos podría encuadrarse la falta de motivación debida al sesgo de automatización por parte del juzgador?

El derecho a la motivación de las resoluciones es la principal garantía del derecho de defensa y se protege constitucionalmente, para el caso de las sentencias, mediante el art. 120.3º CE en relación con el art. 24 CE. Esto tiene una doble función: conocer los razonamientos que determinan el fallo y su control mediante las vías de impugnación y prevenir la arbitrariedad y las resoluciones irrazonadas o irrazonables. Siendo optimistas por disponer de más tiempo que para el recurso de reposición de la vista, si se detecta a tiempo para interponer un recurso de suplicación al amparo del art. 193.a) LRJS por falta de motivación y consiguiente vulneración de la tutela judicial efectiva y provocación de indefensión, para reponer los autos al estado en el que se encontraban en el momento de cometerse la infracción, ¿cómo se graduará a efectos del proceso la gravedad de la actuación acrítica del juzgador que dicta una sentencia que contiene alucinaciones de cualquiera de las diversas clases aquí expuestas?

En cuanto a la nulidad de actuaciones[51]; la falta de motivación fáctica se considera difícilmente subsanable porque quebraría el principio de inmediación y el juzgador que dirige el acto de la vista es el único que puede razonar sobre las pruebas practicadas que valora. Para que la falta de motivación

[49] Para ahondar en ello, véase: Aradas García, A.: «La aportación de una prueba creada con inteligencia artificial en la jurisdicción social». *Anuario da Facultade de Dereito da Universidade de Coruña*, Vol. 28, 2024, pp. 52-64.

[50] Celebi, N., Liu, Q., Karatoprak, M.: «A survey of deep fake detection for trial courts». *arXiv Cornell University preprint arXiv 2205.15792*, 2022.

[51] Olmos Parés, I.: *La nulidad de actuaciones en el proceso social*. Valencia: Tirant lo Blanch, 2018, pp. 1-268.

fáctica provoque la nulidad de la sentencia, tiene que concurrir una arbitrariedad o irracionalidad en la valoración y motivación de la prueba, en cuyo caso la nulidad absoluta de la sentencia acaecería porque caería la base de la misma. La falta de motivación jurídica podría ser subsanada salvo que la sentencia de instancia adolezca seriamente de una motivación suficiente. Por el contrario, no pueden subsanarse las resoluciones que parten de premisas inexistentes o notoriamente erróneas, o que desarrollan argumentos con quiebras lógicas de tal magnitud que las conclusiones no pueden basarse en ninguna de las sazones aducidas. Considerando la exégesis racional que exige el TC[52] para la tutela judicial efectiva, ¿cómo se medirá a efectos del proceso el grado de irracionalidad, arbitrariedad y falta de motivación de un juzgador que cae en el sesgo de automatización?

Y finalmente, lo que debería suscitar mayor inquietud: ¿cómo se impugnará este tipo de negligencia y se logrará depurar el proceso si el juzgador no está obligado a incorporar en la resolución el detalle sobre el uso que ha hecho de la IA?

5. ALGUNAS PROPUESTAS ACOMPAÑANDO A LA DEPURACIÓN NORMATIVA PLANTEADA

Como consideración previa cabe destacar que aunque el CGPJ haya aprobado un protocolo para la gestión de las consultas de la Carrera Judicial sobre la utilización de sistemas de IA y sus usos permitidos y prohibidos, lo cual siempre será positivo, el grado de transparencia sobre los modelos que componen KENDOJ debería ser mayor de inicio, puesto que no se encuentra información sobre si, por ejemplo, se basan en los sistemas RAG previamente conceptualizados (información que necesita analizarse desgranando cada elemento técnico).

5.1. Resoluciones judiciales con un apartado final donde se declare detalladamente si se ha empleado la IA para su formulación

A tenor de todo lo expuesto y en aras de posibilitar vías de impugnación razonables ante un posible empleo negligente de herramientas de IA[53], parece lógico plantear que las personas juezas y magistradas tengan la obligación de incorporar en sus resoluciones un apartado final en el que se detalle si se ha empleado alguna herramienta de IA y en qué medida (obviamente, a partir de ahora únicamente se declarará el empleo de alguna de las herramientas puestas a disposición por parte del CGPJ debido a que de lo contrario estaría admitiendo un incumplimiento de la Instrucción).

Observando todo lo que aquí se ha tratado y siguiendo el razonamiento del cdo. 59 del RIA, el uso de sistemas de IA de alto riesgo por parte de las autoridades garantes del cumplimiento del Derecho no debe derivar en un elemento de desigualdad o exclusión, ni debe ignorarse su impacto y la dificultad para obtener información significativa sobre su funcionamiento, especialmente cuando no existe una transparencia suficiente, lo que imposibilita la impugnación de sus resultados[54].

[52] Por todas, STC (1ª) n. 145/2012 de 2 de julio de 2012 (Rec. 273/2011) F.J.4º.

[53] Teniendo en cuenta que «toda sentencia debe explicitar las razones que le han llevado a un determinado pronunciamiento, lo que permite a las partes conocer el razonamiento subyacente a los efectos de articular eventuales impugnaciones (función endoprocesal) y el propio control de legitimidad democrática de dichas decisiones por parte de la población en general (función extraprocesal)». Juliá Pijoan, M. (2024): «"Las Sentencias actuales no son los mejores datos de entrenamiento para la Inteligencia Artificial». VV.AA.: *La computarización del derecho: a partir del proceso y de los procedimientos judiciales*, Calaza López, S. (Dir.) Madrid, Dykinson, pp. 97-119.

[54] «Dado su papel y su responsabilidad, las actuaciones de las autoridades garantes del cumplimiento del Derecho que implican determinados usos de los sistemas de IA se caracterizan por un importante desequilibrio de poder y pueden tener otros efectos negativos sobre los derechos fundamentales consagrados en la Carta. En particular, si el sistema de IA no está entrenado con datos de buena calidad, no cumple los requisitos adecuados en términos de rendimiento, de precisión o de solidez, o no se diseña y prueba debidamente antes de introducirlo en el mercado o ponerlo en servicio, es posible que señale a personas de manera discriminatoria, incorrecta o injusta. Además, podría impedir el

5.2. Mientras tanto, cabe solicitar el acceso al código fuente al amparo de la jurisprudencia del caso Bosco

La STS (Sala de lo Contencioso-Advo.) núm. 1119/2025 de 11 de septiembre de 2025 [(Rec. 7878/2024) ECLI:ES:TS:2025:3826)] ha supuesto un hito sobre la materia. Primeramente, parte de insistir, sobre la naturaleza del derecho de acceso a la información pública del art. 105.b[55] de la CE, en que, si bien no es absoluto sin perjuicio de su formulación amplia, no se puede limitar por motivos discrecionales de la Administración, sino por la valoración proporcional de los supuestos tasados mediante los arts. 14 y 15 de la Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno (LTAIBG)[56]. El TS se inclina por «interpretar los citados límites de forma restrictiva, a fin de no menoscabar el derecho de acceso, regulado de forma amplia en la Ley» (F.D. 6º.B)[57]. Al mismo tiempo, entre otras cuestiones, se subraya que la configuración del derecho de acceso, si bien no tiene la categoría formal de fundamental (aunque siguiendo la pauta interpretativa del art. 10. 2 CE en relación con el art. 42 de la CDFUE[58] puede considerarse como tal), se encuentra relacionado con, entre otros, el derecho fundamental a la tutela judicial efectiva del art. 24 CE y el derecho fundamental a la seguridad jurídica del art. 9.3 CE, que es el que más interesa a efectos de lo que aquí se plantea. El precepto 105.b CE contiene el principio objetivo rector de la actuación de la Administración que emana de las exigencias democráticas y de transparencia, constituyendo a su vez un derecho subjetivo ejercitable ante las AA.PP. en los términos desarrollados por la LTAIBG.

Sentados los elementos esenciales de la doctrina sobre el derecho de acceso, lo novedoso de esta Sentencia es su valoración respecto a su aplicación en el ámbito de los sistemas o aplicaciones informáticas de las AA.PP. A efectos del caso que está siendo objeto de análisis, una de las dudas que puede suscitar la interpretación de esta jurisprudencia que posibilita el acceso al código fuente es que, mientras en el caso BOSCO se trataba de un sistema plenamente automatizado que resolvía las solicitudes del bono social eléctrico, el escenario que aquí se plantea es el de sistemas de apoyo para la toma de decisiones judiciales y sistemas semi automatizados para la recuperación de información de las bases de datos. El TS fundamenta que «como es lógico, la explicabilidad de las aplicaciones informáticas, así como de los algoritmos que las sustentan, utilizadas por las Administraciones públicas, es objeto de una creciente demanda ciudadana, como condición inexcusable para preservar la rendición de cuentas y la fiscalización de las decisiones de los poderes públicos y, en último término, como garantía efectiva frente a la arbitrariedad o los sesgos discriminatorios en la toma de decisiones total o parcialmente automatizadas». En teoría, en el ámbito de las decisiones judiciales no debería concurrir una automatización plena, pero una mala *praxis* debida al sesgo de automatización nos coloca en ese escenario de las actuaciones viciadas que pueden pasar desapercibidas por las dinámicas en las que nos encontramos inmersos, que es como se ha iniciado la exposición del problema objeto de estudio.

ejercicio de importantes derechos procesales fundamentales, como el derecho a la tutela judicial efectiva y a un juez imparcial, así como el derecho a la defensa… sobre todo cuando dichos sistemas de IA no sean lo suficientemente transparentes y explicables ni estén suficientemente bien documentados».

[55] «El acceso de los ciudadanos a los archivos y registros administrativos, salvo en lo que afecte a la seguridad y defensa del Estado, la averiguación de los delitos y la intimidad de las personas».

[56] «BOE» n. 295, de 10/12/2013.

[57] Reiterada doctrina desde la STS 579/2024 de 8 de abril de 2024 (rec. 681/2022), FJ 4º. ECLI: ES:TS:2024:2249.

[58] «Todo ciudadano de la Unión y toda persona física o jurídica que resida o tenga su domicilio social en un Estado miembro tiene derecho a acceder a los documentos de las instituciones, órganos y organismos de la Unión, cualquiera que sea su soporte».

Si se reconoce el principio de transparencia algorítmica y el derecho de acceso a la ciudadanía, incluso sin necesidad de motivación[59], parece legítimo que las partes procesales puedan conocer la arquitectura de los sistemas de IA que utiliza el juzgador. Uno de los reconocimientos fundamentales del TS es que el derecho de acceso resulta especialmente relevante por los riesgos que derivan del empleo de las nuevas tecnologías por parte de los poderes públicos (en aquel caso, dada la especial garantía que merece el ámbito del reconocimiento de derechos sociales a colectivos en situación de vulnerabilidad basado en la toma de decisiones automatizadas). En el caso que aquí interesa también se trata de velar por la integridad del proceso, que puede quedar viciada, no por una automatización en sentido estricto, sino por el sesgo de automatización. El reconocimiento del principio de transparencia algorítmica impone a las AA.PP. la obligación de facilitar a la ciudadanía la información fundamental sobre los algoritmos que emplean para la toma de decisiones o su código fuente, entendiéndose como una de las traducciones del principio de transparencia establecido por el citado art. 105.b) de la CE. El TS incluso va más allá afirmando que «evitar la opacidad del algoritmo o el código fuente se muestra consustancial al Estado democrático de Derecho, sujeto al principio de transparencia, y refuerza la confianza ciudadana en el correcto funcionamiento de las Administraciones públicas» y, de hecho, la seguridad jurídica reviste mayor importancia que dicha confianza.

Dicho deber no se puede tomar como algo programático, sino como una fórmula que exige la obligación a las AA.PP. de garantizar la transparencia y el acceso a la información pública[60]. Al mismo tiempo, el TS fundamenta con claridad que el « el mero riesgo de eventuales perjuicios para el derecho de propiedad intelectual de las Administración pública, con motivo de su uso o explotación no autorizada, como consecuencia del acceso al código fuente, por sí solo, no puede constituir causa de exclusión del derecho de acceso». En este caso, no sabríamos si concurre tal perjuicio porque parece que los modelos KENDOJ se han desarrollado internamente y no son una adquisición externa. La limitación del derecho de acceso solo se justificaría tras una ponderación de los intereses concurrentes. Por el momento, se plantea un control por parte del CGPJ, pero eso no posibilita que las partes puedan conocer la existencia de una mala *praxis* y puedan impugnar. Como no podía ser de otra manera, el TS también reconoce que el acceso puede contribuir a la mejora del código por someterlo al escrutinio de diversos actores. Razón por la cual, a modo preventivo cabe insistir en la necesidad de emplear los *sandboxes* regulatorios interdisciplinares.

5.3. ¿Podría automatizarse un sistema pericial que revise toda la prueba obrante en autos junto a los escritos y las grabaciones de las vistas?

La actual digitalización del proceso parece una vía viable para desarrollar procesos de peritaje automático de la prueba obrante en los expedientes, junto a las vistas y los escritos.

[59] Desde la STS n. 1519/2020 de 12 de noviembre de 2020 (Rec. 5239/2019), FJ 4º.7. ECLI:ES:TS:2020:3870, a diferencia de los requisitos que fijó el TEDH en su Sentencia 18030/11, de 8 de noviembre de 2016 (asunto Maygar Helsinki Bizottság c. Hungría): («i) propósito del acceso —la finalidad buscada debe generar debate público o contribuir a la labor de vigilancia de los poderes públicos—; (ii) naturaleza de la información -la información debe ser de interés público, es decir, que proporcione transparencia sobre los asuntos públicos o sea relevante para la sociedad en su conjunto-; (iii) rol del solicitante -debe tratarse de un solicitante que desempeñe funciones de vigilancia social asociadas a la guardia y custodia de la democracia (el TEDH entiende que cumplen estas condiciones los medios de comunicación y periodistas, pero también organizaciones no gubernamentales, investigadores académicos, escritores sobre materias de interés público, blogueros, influencers, etc.)-, y (iv) existencia y disposición de la información -la información debe estar disponible, sin que sea necesario acometer una labor de recopilación». Esto es, la LTAIBG reconoce el derecho de acceso a la información pública con mayor amplitud que el TEDH.

[60] Por todas, STS n. 1752/2022, de 23 de diciembre de 2022 (Rec. 1763/2021), F.J. 5º. ECLI:ES:TS:2022:4891.

5.4. ¿Los Ilustres Colegios de la Abogacía deberían contribuir a democratizar el acceso a herramientas de IA especializadas?

En la línea de lo propuesto por la Relatora, facilitar el acceso a herramientas de IA especializadas, subvencionadas o gratuitas, para democratizarlas y disminuir la desigualdad de medios parece una iniciativa razonable. No obstante, la desigualdad no solo radica en el acceso a las herramientas, porque en palabras de la Relatora «los actores pudientes que intervienen con frecuencia en las actuaciones procesales, principalmente las empresas, con acceso a enormes volúmenes de datos sobre acuerdos privados que pueden utilizarse para entrenar IA especializada, obtienen ventajas particulares. Los actores que intervienen con frecuencia en las actuaciones procesales pueden incluso salir ganando en relación con los sistemas de IA utilizados para la administración judicial». Tras desgranar la arquitectura de los sistemas, resulta plenamente lógico que exista esa dinámica. En consecuencia, también deberá depurarse otra clase de sesgo, que deriva de la desigualdad estructural de medios procesales.

6. CONCLUSIONES

Aunque las conclusiones se han ido exponiendo a lo largo del artículo, a continuación se insiste en algunas de ellas.

El surgimiento de conductas de mala *praxis* por parte de los operadores jurídicos como consecuencia del sesgo de automatización exige un exhaustivo esfuerzo de alfabetización técnico e interdisciplinar, que visibilice la diversidad de errores y alucinaciones específicos del ámbito jurídico y procesal, más allá de la casuística más burda.

Resulta inadmisible adoptar una posición pasiva y conformista. Por un lado, los sistemas RAG son prometedores, pero todavía obligan a una profunda revisión y un entrenamiento técnico-jurídico constante, que puede no ser garantista y desvirtuar la eficiencia que se busca. También procede reflexionar sobre el poder epistémico, tanto desde la perspectiva de la economía política que subyace a los modelos de IA, como desde las fórmulas que no cuestionan la diversidad e idoneidad de los datos que se utilizan para nutrirlos, especialmente en lo que a la jurisprudencia se refiere.

Por otro lado, se detectan problemas en la redacción de la Instrucción del CGPJ: 1) con la expresión «pueda ser libre y enteramente modificado» del ordinal 7º.4 en relación con el 6º.3.c) y el 6º.4. se faculta a los juzgadores a validar como resolución judicial o procesal un borrador generado con IAG (con el riesgo que supone, aunque intenten revisarlo, por todo lo evidenciado); 2) afirma que los borradores generados no tendrán, en ningún caso, la consideración de decisiones automatizadas (lo cual carecería de sentido ante una mala *praxis* del juzgador viciado por el sesgo de automatización) [ord. 7.3]; 3) establece un régimen sancionador indeterminado y una dinámica reactiva en lugar de preventiva [ord. 11º]; 4) deja la formación y capacitación en el plano de la voluntariedad [ord. 13º], ignorando algunas de las exigencias del RIA (entre otros, el art. 26.2).

En esencia, este estudio conduce a responder las dos preguntas que se plantean de forma negativa. En la jurisdicción social, observando transversalmente todos los problemas técnicos y procesales descritos, si no se toman medidas suficientes, los principios vertebradores del proceso no serán sostenibles ante el uso negligente de la IA por la contraparte o el juzgador, porque la celeridad de los hitos procesales determinantes, junto a la ausencia de medios técnicos y márgenes temporales suficientes para la detección y depuración de un proceso que arrastre vicios en cualquiera de las casuísticas planteadas, provoca una inseguridad jurídica gravísima.

Se considera que ante un caso de sesgo de automatización por parte del juzgador, la mejora de la tutela judicial efectiva y la defensa pueden facilitarse mediante la preceptiva inclusión de un nuevo apartado en las resoluciones, por el que se declare y detalle si se ha utilizado la IA y cómo. Por el momento, cabe ampararse en la jurisprudencia del caso Bosco y pensar en nuevas formas técnicas de controlar la veracidad de todo lo obrante en autos.

7. BIBLIOGRAFÍA

Aradas García, A.: «La aportación de una prueba creada con inteligencia artificial en la jurisdicción social». *Anuario da Facultade de Dereito da Universidade de Coruña*, Vol. 28, 2024, pp. 52-64.

Banks, I.: "Judges-in-the-loop? Judicial involvement in human oversight of high-risk decision support systems under the EU AI Act. International" *Journal of Law and Information Technology, Oxford University Press*, 2026, Vol. 34, n. 1, pp.1-23.

Brenan-Márquez, K. Levy, K. y Susser, D.: «Strange Loops: Apparent versus actual human involvement in automated decision making». *Berkeley Technology Law Journal,* Vol. 34, 2019, pp 745–772.

Celebi, N., Liu, Q., Karatoprak, M.: «A survey of deep fake detection for trial courts». *arXiv Cornell University preprint* arXiv 2205.15792, 2022, pp 227–238

Dantart, a.: «Inteligencia Artificial jurídica y el desafío de la veracidad: análisis de alucinaciones, optimización de RAG y principios para una integración responsable». *arXive-prints*, *arXiv:2509.09467*, *Cornell University*, 2025, pp. 1-118. https://arxiv.org/abs/2509.09467

Deroy, A., Ghosh, K., Ghosh, S.: «How ready are pre-trained abstractive models and LLMs for legal case judgement summarization?». *arXiv preprint arXiv:2306.01248 Cornell University*, 2023.

Kalai, A.T., et al.: «Why language models hallucinate». *arXiv preprint Cornell University, arXiv:2509.04664*, 2025, pp. 1-36.

Juliá Pijoan, M. (2024): «Las Sentencias actuales no son los mejores datos de entrenamiento para la Inteligencia Artificial». VV.AA.: *La computarización del derecho: a partir del proceso y de los procedimientos judiciales*, Calaza López, S. (Dir.) Madrid, Dykinson.

Magesh, V., *et al*: "Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools". *Journal of empirical legal studies*, 2025, vol. 22, n.º 2, pp. 216-242.

Nores Torres, L.: «El orden social y el proceso laboral tras la LO 1/2025, de 2 de enero». *Trabajo y Empresa. Revista de Derecho del Trabajo* Vol. 4, n. 1. 2025, pp. 39-62

Olmos Parés, I.: *La nulidad de actuaciones en el proceso social*. Valencia: Tirant lo Blanch, 2018, pp. 1-268.

Peralta Gutierrez, A.: «Impacto comparado de la IA en la Administración de Justicia. La Guía Práctica de la Inteligencia Artificial». Nº 5, 2025. *Aranzadi La Ley* 1–32.

Satterthwaite, M.: *Informe sobre la Independencia de los magistrados y abogados*. AGNU, 2025, pp. 1-23.

Tascón López, R.: «Algunas reflexiones sobre la situación del proceso social tras la Ley Orgánica 1/2025, 2 de enero, de medidas en materia de eficiencia del servicio público de justicia». *Revista Crítica de Relaciones de Trabajo, Laborum*. n 15º, vol. 2º, 2025, pp 29–52.