

DESAFÍOS DE LA IA GENERATIVA EN DOMINIOS DE ALTO RIESGO: REQUISITOS PARA LA SIMPLIFICACIÓN JUDICIAL FIABLE

CHALLENGES OF GENERATIVE AI IN HIGH-RISK DOMAINS: REQUIREMENTS FOR RELIABLE JUDICIAL SIMPLIFICATION

Ana-Belén GIL-GONZÁLEZ¹, Silvia DOMÍNGUEZ CASTRO², Clara RODRÍGUEZ MARCOS³,
Alberto JIMÉNEZ SÁNCHEZ⁴, Javier MANSO CONDE⁵ y Jorge Enrique GONZÁLEZ GONZALO⁶

¹ *Universidad de Salamanca. Salamanca, Castilla y León, España*

✉ abg@usal.es

🌐 <https://orcid.org/0000-0001-7235-6151>

² *Universidad de Salamanca. Salamanca, Castilla y León, España*

✉ sdominguez@air-institute.com

🌐 <https://orcid.org/0009-0006-5290-7394>

³ *Universidad de Salamanca. Salamanca, Castilla y León, España*

✉ crodriguez@air-institute.com

🌐 <https://orcid.org/0009-0005-2713-9269>

⁴ *Universidad de Salamanca. Salamanca, Castilla y León, España*

✉ ajimenez@air-institute.com

🌐 <https://orcid.org/0009-0004-4183-0008>

⁵ *Universidad de Salamanca. Salamanca, Castilla y León, España*

✉ jmanso@air-institute.com

🌐 <https://orcid.org/0009-0005-5187-9566>

⁶ *Universidad de Salamanca. Salamanca, Castilla y León, España*

✉ jgonzalez@air-institute.com

🌐 <https://orcid.org/0009-0004-4053-3503>

RESUMEN: El acceso a la justicia requiere no solo la posibilidad de acudir a un tribunal, sino la capacidad de comprender sus decisiones. En España, la Ley 6/2022 establece la accesibilidad cognitiva como un derecho, promoviendo el uso de la lectura fácil (LF). Sin embargo, la saturación del sistema judicial hace inviable la adaptación manual masiva de sentencias. La inteligencia artificial generativa (GenIA) surge como una solución potencial,

pero su aplicación en justicia es clasificada como de «alto riesgo» por el Reglamento de IA de la UE debido a problemas como las alucinaciones y la falta de explicabilidad. Este trabajo analiza dichos desafíos técnicos y jurídicos, proponiendo una arquitectura de fiabilidad basada en tres requisitos innegociables: anclaje factual con trazabilidad documental, supervisión humana obligatoria y segregación funcional de roles. El objetivo es definir un estándar que permita automatizar la simplificación del lenguaje jurídico sin comprometer el rigor legal ni las garantías procesales.

Palabras clave: inteligencia artificial generativa; lectura fácil; acceso a la justicia.

ABSTRACT: Access to justice requires not only the possibility of going to court, but also the ability to understand its decisions. In Spain, Law 6/2022 establishes cognitive accessibility as a right, promoting the use of Easy Reading (LF). However, the saturation of the judicial system makes the manual adaptation of judgments unfeasible. Generative Artificial Intelligence (GenAI) emerges as a potential solution, but its application in justice is classified as «High Risk» by the EU AI Regulation due to problems such as hallucinations and lack of explainability. This paper analyses these technical and legal challenges, proposing a reliability architecture based on three non-negotiable requirements: factual anchoring with documentary traceability, mandatory human supervision and functional segregation of roles. The aim is to define a standard that allows the simplification of legal language to be automated without compromising legal rigour or procedural guarantees.

Keywords: generative artificial intelligence; easy reading; access to justice.

1 INTRODUCCIÓN

El acceso a la justicia no consiste solo en la posibilidad de acudir a un tribunal, sino en la capacidad de comprender las decisiones que allí se toman. Sin embargo, el lenguaje jurídico actual suele ser oscuro, excesivamente técnico y difícil de entender para el ciudadano medio. Esta complejidad crea una barrera de comunicación que afecta a los colectivos más vulnerables, limitando su autonomía para entender sus propios derechos y obligaciones [1, 18].

En España, la legislación ha dado un paso decisivo para resolver este problema. La aprobación de la Ley 6/2022 [12] ha convertido la accesibilidad cognitiva en un derecho exigible, obligando a las Administraciones públicas a adaptar sus comunicaciones para que sean comprensibles. Esto incluye el uso de la lectura fácil (LF), una técnica de redacción y diseño regulada por la norma UNE 153101:2018 [2]. No obstante, la aplicación de esta normativa se enfrenta a un obstáculo: el sistema judicial está saturado y la adaptación manual de cada sentencia es una tarea inviable debido al enorme volumen de documentos que se generan a diario [6].

Ante este escenario, la inteligencia artificial generativa (GenIA) se presenta como una herramienta capaz de realizar esta tarea. Sin embargo, el uso de modelos de lenguaje estándar (como los utilizados en ChatGPT) conlleva riesgos en el ámbito legal, debido a su tendencia a inventar información (las llamadas «alucinaciones») y a la falta de transparencia sobre cómo

toman decisiones [3]. Por este motivo, el reciente Reglamento de Inteligencia Artificial de la Unión Europea [11] clasifica el uso de la inteligencia artificial (IA) en la justicia como un sistema de «alto riesgo», exigiendo garantías estrictas de supervisión y fiabilidad.

El objetivo de este trabajo es analizar estos desafíos y definir qué requisitos técnicos debe cumplir un sistema de IA para automatizar la LF sin entrar en conflicto con el rigor legal.

2. EL IMPERATIVO NORMATIVO: DE LA RECOMENDACIÓN A LA OBLIGACIÓN

Durante décadas, la simplificación del lenguaje administrativo en España se mantuvo en el terreno de las recomendaciones de estilo o los códigos de buenas prácticas [1]. Sin embargo, el marco jurídico ha sufrido una transformación radical donde el punto de inflexión fue la Ley 6/2022, de 31 de marzo, que modifica el Texto Refundido de la Ley General de derechos de las personas con discapacidad [12]. Esta norma es trascendental porque define y regula, por primera vez, la accesibilidad cognitiva dentro del ordenamiento jurídico español, estableciendo que la Administración de justicia debe garantizar que sus procesos sean comprensibles para todas las personas, independientemente de sus capacidades intelectuales [4, 24]. Ya estamos ante un derecho subjetivo exigible ante los tribunales, cuyo incumplimiento puede derivar en la nulidad de actuaciones por indefensión.

La accesibilidad se ha tecnificado a través de estándares como la norma UNE 153101:2018 [2], la cual aclara una confusión habitual en la práctica forense: la LF no es un mero resumen ni una versión «infantilizada» del texto original [5]. Se trata de una técnica de redacción y publicación que impone reglas estrictas, tales como la limitación de la longitud de las oraciones, el uso de glosarios laterales para términos complejos y la eliminación de metáforas o lenguaje figurado que puedan distorsionar el mensaje [2, 19]. La norma subraya que la adaptación no es solo lingüística, sino también visual, exigiendo tipografías específicas, márgenes amplios y una jerarquía de la información que facilite el rastreo visual, elementos que la Ley 6/2022 ha convertido en condiciones básicas de accesibilidad [12, 21].

En este contexto, la adaptación documental converge con la disciplina académica del *Legal Design* (Diseño Legal). Este enfoque, teorizado por investigadoras como Margaret Hagan [10], señala que la justicia debe ser diseñada «centrada en el usuario» para ser efectiva. Aplicado a las sentencias, el *Legal Design* implica que la simplificación no se limita a reescribir párrafos, sino a reestructurar la información para hacerla navegable y reducir la carga cognitiva [9]. Estudios recientes demuestran que la combinación de un lenguaje claro y de un diseño visual mejora la capacidad de las personas para tomar decisiones legales informadas [13]. Por consiguiente, cualquier sistema de IA no puede limitarse a ser un «traductor» de palabras; debe operar con una interoperabilidad técnica y procedimental que le permita actuar como un diseñador de información, organizando visualmente el contenido para garantizar la comprensión [20, 23].

3. RIESGOS DE LA IA GENERATIVA EN DOMINIOS DE ALTO RIESGO

La integración de grandes modelos de lenguaje (LLM) en la Administración Pública no puede realizarse de manera acrítica. El ámbito jurídico exige un estándar de precisión factual y seguridad normativa que los modelos generativos actuales no pueden garantizar [25]. Esta incompatibilidad ha motivado que el legislador europeo, en el reciente Reglamento (UE) 2024/1689 [17], clasifique explícitamente en su Anexo III los sistemas de IA destinados a la Administración de Justicia como sistemas de «alto riesgo» [16]. Esta clasificación no es arbitraria, sino que responde a tres amenazas sistémicas que comprometen las garantías del debido proceso.

En primer lugar, la naturaleza de los LLM da lugar al fenómeno de las «alucinaciones». Tal como advierten Bender *et al.* [3], estos modelos no acceden a una base de datos de verdad, sino que operan prediciendo la siguiente palabra más probable basándose en patrones estadísticos, sin conexión con la realidad. En el dominio legal, estudios recientes realizados por la Universidad de Stanford han demostrado que los modelos de propósito general (como GPT-4) inventan jurisprudencia o citan normativas inexistentes en un porcentaje alarmante de las consultas jurídicas complejas (entre el 15 % y el 25 %), produciendo una veracidad aparente, pero factualmente falsa [8]. En una sentencia de LF, una alucinación que altere un hecho probado o la duración de una pena no es un error informático, sino una vulneración directa de la tutela judicial efectiva.

En segundo lugar, la opacidad algorítmica o el problema de la «Caja Negra» contraviene el principio de motivación de las resoluciones judiciales. El artículo 24 de la Constitución española y la normativa europea exigen que cualquier decisión administrativa sea explicable y trazable [7]. Sin embargo, el funcionamiento interno de las redes neuronales profundas es ininteligible para el operador humano, lo que impide auditar por qué el sistema decidió simplificar un término de una manera y no de otra. El Grupo de Expertos de Alto Nivel de la UE [11] ha señalado que esta falta de explicabilidad es inaceptable en sectores donde se dirimen derechos fundamentales, exigiendo mecanismos de supervisión humana obligatorios.

Finalmente, existe el riesgo de la sobresimplificación, debido a que el léxico jurídico posee una carga semántica precisa donde la eliminación de matices altera la sustancia del mensaje. Investigaciones sobre *Legal Tech* advierten que los modelos entrenados para maximizar la legibilidad tienden a eliminar cláusulas subordinadas o términos técnicos que son necesarios en una sentencia [22]. Una IA que priorice la facilidad de lectura por encima del rigor legal podría producir textos muy accesibles, pero jurídicamente vacíos o incorrectos, despojando al ciudadano de la información necesaria para su defensa y creando una falsa sensación de comprensión [15].

4. REQUISITOS PARA UNA IA DE SIMPLIFICACIÓN FIABLE

Con el objetivo de que la automatización de la LF sea compatible con el Estado de derecho, cualquier arquitectura propuesta debe satisfacer, como mínimo, tres condiciones de fiabilidad que mitiguen los riesgos inherentes a la GenIA.

4.1. REQUISITO 1: ANCLAJE FACTUAL Y TRAZABILIDAD DOCUMENTAL

El principio de veracidad judicial exige que la IA abandone su comportamiento creativo. El modelo no debe generar texto basándose en su entrenamiento previo (que puede contener sesgos), sino exclusivamente en la información contenida en la sentencia original. Técnicamente, esto requiere mecanismos de recuperación de información que obliguen al algoritmo a citar o referenciar el fragmento exacto del documento fuente que justifica cada frase simplificada [14]. Si el sistema no puede trazar el origen de una afirmación hasta el texto original, no debe generarla. Solo mediante esta trazabilidad documental es posible realizar la auditoría *ex post* que exige el Reglamento de IA [17].

4.2. REQUISITO 2: SUPERVISIÓN HUMANA

La tecnología debe concebirse como una herramienta de asistencia, nunca de sustitución. Siguiendo las directrices de la OCDE sobre IA en la justicia y el artículo 14 del *AI Act*, la decisión final debe recaer siempre en un operador humano [16]. La IA debe actuar como un «borrador avanzado», automatizando la tarea de reescritura, pero dejando el control de calidad en manos del humano. Para que esta supervisión sea real, la interfaz del sistema debe resaltar los cambios realizados, permitiendo al supervisor verificar que la simplificación léxica no ha alterado la calificación jurídica de los hechos.

4.3. REQUISITO 3: SEGREGACIÓN FUNCIONAL

La metodología humana de la LF se basa en la separación de roles: un profesional adapta el texto y otro verifica su accesibilidad [2]. La IA debe emular esta arquitectura de control. No es suficiente con pedirle al modelo que «simplifique»; es necesario disociar el proceso en dos etapas: una instancia encargada de la redacción en lenguaje claro y una segunda instancia, independiente y con instrucciones adversarias, encargada de la validación normativa [26]. Esta segregación, propia de los sistemas de alta fiabilidad, reduce la tasa de error lógico, asegurando que la búsqueda de la sencillez no comprometa el rigor legal.

5. CONCLUSIÓN

La irrupción de la IA en el sector público nos sitúa ante un dilema: adoptar herramientas opacas que sacrifican garantías por eficiencia o construir sistemas que fortalezcan el acceso a la justicia. El análisis realizado evidencia que la adaptación a lectura fácil, exigida por la Ley 6/2022 [12], es inabordable manualmente, pero peligrosa si se delega en modelos de «Caja Negra». Para que la tecnología cumpla su promesa democratizadora, es necesario transitar hacia arquitecturas supervisadas que integren el anclaje factual y la validación humana como requisitos innegociables. Solo así garantizaremos que el derecho a entender de la ciudadanía se materialice sin socavar la seguridad jurídica que sustenta nuestra democracia.

6. AGRADECIMIENTOS

Esta investigación forma parte del Proyecto de Cátedra Internacional sobre Inteligencia Artificial Fiable y Reto Demográfico dentro de la Estrategia Nacional de Inteligencia Artificial (ENIA), en el marco del Plan Europeo de Recuperación, Transformación y Resiliencia. Referencia: TSI-100933-2023-0001. Este proyecto está financiado por la Secretaría de Estado para la Digitalización y la Inteligencia Artificial y por la Unión Europea (Next Generation).

7. REFERENCIAS

1. Alcaraz Varó, E., Hughes, B., & Gómez Martínez, M. A. (2014). *El español jurídico*. Ariel.
2. Asociación Española de Normalización. (2018). *Lectura fácil: Pautas y recomendaciones para la elaboración de documentos* (Norma UNE 153101:2018 EX). AENOR.
3. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM FAccT Conference on Fairness, Accountability, and Transparency* (pp. 610-623). ACM. <https://doi.org/10.1145/3442188.3445922>
4. Burgos Martínez, M. (2025). La paradoja de la aplicación del derecho a comprender en España: Una revisión normativa. *Sphera Publica*, 1(25), 112-141.
5. Centro de Referencia Estatal de Autonomía Personal y Ayudas Técnicas. (2023). *Lectura fácil (Ficha 3)* (Serie 10 fichas sobre la accesibilidad cognitiva por derecho). CEAPAT.
6. Consejo General del Poder Judicial. (2024). *Memoria anual 2024 sobre el estado, funcionamiento y actividades de los Juzgados y Tribunales*. CGPJ.
7. Cotino Hueso, L. (2023). Qué concreta transparencia e información de algoritmos e inteligencia artificial es la debida. *Revista Española de la Transparencia*, 16, 17-63. <https://doi.org/10.51915/ret.272>
8. Dahl, M., Maughan, J., Glazer, L., & Nyarko, J. (2024). Large legal fictions: Profiling legal hallucinations in large language models. *Journal of Legal Analysis*, 16(1), 64-93. <https://doi.org/10.1093/jla/lae003>
9. Hagan, M. (2018). A human-centered design approach to access to justice: Generating new prototypes and hypotheses for intervention to make courts user-friendly. *Indiana Journal of Law and Social Equality*, 6(2), 199-239.

10. Hagan, M. (2019). Participatory design for innovation in access to justice. *Daedalus*, 148(1), 120-127. https://doi.org/10.1162/daed_a_00544
11. High-Level Expert Group on AI. (2019). *Ethics guidelines for trustworthy AI*. European Commission.
12. Jefatura del Estado. (2022, 1 de abril). Ley 6/2022, de 31 de marzo, de modificación del Texto Refundido de la Ley General de derechos de las personas con discapacidad y de su inclusión social, aprobado por el Real Decreto Legislativo 1/2013, de 29 de noviembre, para establecer y regular la accesibilidad cognitiva y sus condiciones de exigencia y aplicación. *Boletín Oficial del Estado*, (78), 43391-43401.
13. Laskowski, C., Rhee, S., & Shadiac, J. (2024). *Generative artificial intelligence and access to justice: Possibilities, concerns, best practices*. National Center for State Courts.
14. Lewis, P., Perez, E., Piktus, A., Petroni, F., Lewis, V., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33, 9459-9474.
15. Montolío Durán, E. (Ed.). (2012). *Hacia la modernización del discurso jurídico*. Universitat de Barcelona.
16. OECD. (2019). *Equal access to justice for inclusive growth*. OECD Publishing. <https://doi.org/10.1787/597f5b7f-en>
17. Parlamento Europeo, & Consejo de la Unión Europea. (2024). Reglamento (UE) 2024/1689, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial (Reglamento de Inteligencia Artificial). *Diario Oficial de la Unión Europea*, L, 2024/1689.
18. Plena Inclusión España. (2020). *10 consejos para mejorar el acceso a la justicia: Adaptación a lectura fácil de las directrices internacionales de la ONU*.
19. Plena Inclusión España. (2018). *Norma UNE 153101:2018 EX de lectura fácil: Pautas y recomendaciones para la elaboración de documentos*. Recuperado el 17 de diciembre de 2025, de <https://www.plenainclusion.org/publicaciones/buscador/norma-une-1531012018-ex-de-lectura-facil-pautas-y-recomendaciones-para-la-elaboracion-de-documentos/>
20. PwC Tax & Legal. (2022). *El derecho a entender: La comunicación clara como palanca de transformación jurídica*.
21. Real Patronato sobre Discapacidad. (2023). *España País Accesible: II Plan Nacional de Accesibilidad Universal 2023-2032*. Ministerio de Derechos Sociales y Agenda 2030.
22. Saggion, H. (2017). *Automatic text simplification*. Morgan & Claypool Publishers. <https://doi.org/10.2200/S00700ED1V01Y201602HLT032>
23. Simshaw, D. (2025). Interoperable legal AI for access to justice. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.5425095>
24. Tu Web Accesible. (2022). *Ley 6/2022: La nueva accesibilidad cognitiva en España*. Recuperado el 17 de diciembre de 2025, de <https://tuwebaccesible.es/ley-6-2022-accesibilidad-cognitiva/>
25. Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, M., Balle, B., Kasoor, S., Collins, Z., Burns, L., Hendricks, L. A., Wheeler, T., Isaac, W., & Gabriel, I. (2022). Taxonomy of risks posed by language models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 214-229). ACM. <https://doi.org/10.1145/3531146.3533088>
26. Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Chang, B., Wang, Q., Jin, J., Fu, H., & Zhang, Y. (2023). *The rise and potential of large language model based agents: A survey* (arXiv:2309.07864). arXiv. <https://doi.org/10.48550/arXiv.2309.07864>