

Más allá de la caja negra: la IA en los juzgados del orden civil (una propuesta heurística)

Cristina Caja Moya

Universidad del Atlántico Medio

Elio Quiroga Rodríguez

Universidad del Atlántico Medio

Fecha de presentación: enero 2025

Fecha de aceptación: mayo 2025

Fecha de publicación: octubre 2025

Resumen

El artículo explora la potencial implementación de modelos de inteligencia artificial basados en redes neuronales supervisadas en el ámbito de la jurisdicción civil, analizando detalladamente las posibilidades y desafíos de introducir tecnologías avanzadas como los modelos transformadores en el sistema judicial. Se propone una metodología experimental que permitiría evaluar la fiabilidad, precisión y capacidad explicativa de estos modelos, utilizando como campo de pruebas los juzgados civiles españoles, aprovechando la estructura jerárquica y predecible de su legislación, la amplia disponibilidad de datos históricos etiquetados y la relativa baja complejidad ética de ciertos procedimientos civiles en comparación con otras jurisdicciones. El estudio profundiza en las características técnicas que hacen de los modelos como BERT, herramientas prometedoras para asistir a los profesionales jurídicos, destacando especialmente su capacidad para comprender contextos legales, identificar patrones en la jurisprudencia y ofrecer explicaciones transparentes de sus decisiones.

Palabras clave

inteligencia artificial; redes neuronales supervisadas; jurisdicción civil; modelos transformadores; explicabilidad de IA; tecnología jurídica

Beyond the black box: AI in civil order courts (a heuristic proposal)

Abstract

The article examines the potential implementation of artificial intelligence models based on supervised neural networks within civil jurisdiction, providing a detailed overview of the possibilities and challenges associated with integrating advanced technologies, such as transformative models, into the judicial system. An experimental methodology is suggested to evaluate the reliability, accuracy, and explanatory capacity of these models, using Spanish civil courts as a basis for evidence. This approach benefits from the hierarchical and predictable nature of its legislation, the extensive availability of historical labelled data, and the relatively low ethical complexity of certain civil procedures compared to other jurisdictions. The study further explores the technical characteristics that render models like BERT promising tools for aiding legal professionals, particularly emphasizing their capability to comprehend legal contexts, identify patterns in case law, and offer transparent explanations for their decisions.

Keywords

artificial intelligence; supervised neural networks; civil jurisdiction; transformative models; AI explainability; legal technology

Introducción y metodología

El avance en la inteligencia artificial (IA) ha abierto nuevas posibilidades de forma inusitada en diversos campos, incluido el ámbito jurídico, que autores de prestigio como Jordi Nieva Fenoll no han dudado en calificar como «cambio generacional» (Nieva Fenoll, 2022). Todo ello está llevando a considerar, si bien por ahora teóricamente, la implementación y evaluación de IA diseñadas para participar en juicios, operando bajo la supervisión de profesionales humanos, o siendo autónomas.

Desde la perspectiva de la filosofía del derecho, el uso de tecnologías basadas en IA en la toma de decisiones judiciales plantea desafíos y oportunidades (Torrado, 2022). En términos de la teoría de la justicia, estas herramientas podrían contribuir a garantizar una mayor consistencia y predictibilidad en las resoluciones judiciales, reduciendo la arbitrariedad y el impacto de los sesgos humanos, entrando en el terreno de la llamada «justicia procedimental» (Cores, 2023). Sin embargo, también generan preguntas fundamentales sobre la propia naturaleza de la justicia: ¿puede una decisión automatizada, basada en patrones y datos históricos, captar la dimensión ética, contextual y humana que caracteriza el trabajo judicial?

Dentro del marco del positivismo jurídico, podría argumentarse que la aplicación rigurosa de normas a través de estas tecnologías se alinea con una concepción formalista del derecho (Moreso, 2017). No obstante, perspectivas más

realistas o integradoras destacan la importancia de la interpretación, la equidad y la sensibilidad hacia las particularidades de cada caso, elementos que estas herramientas aún no pueden abordar plenamente, lo que llevaría al término conocido como *defeasibility*, cuando un sistema de IA no es capaz de introducir un elemento de adaptabilidad en el razonamiento jurídico, reconociendo que el derecho no puede ser un sistema completamente cerrado o inflexible (Prakken, 2004). Por lo tanto, estas tecnologías *a priori* no parece que estuvieran en condiciones, al menos en su estado actual, de reemplazar a juez alguno, sino de actuar como instrumentos de apoyo que enriquezcan el razonamiento jurídico, garantizando al mismo tiempo la transparencia y el control humano en el proceso de toma de decisiones; todo ello podría invitar, y así lo plantean los autores, a considerar que explorar las posibilidades de estas tecnologías, especialmente los *large language models* (LLM) o IA generativas de lenguaje, desde el punto de vista experimental, podrían llevar a elementos de decisión que hasta ahora no han sido apreciados. Estas han demostrado una versatilidad realmente sorprendente, en los más diversos campos (Xu *et al.*, 2021).

1. La elección del modelo experimental

Un conocido problema es que ciertos modelos de LLM, y concretamente de IA generativa de lenguaje (la que sería

más útil en estos casos), no garantizan, por definición de su diseño, la repetibilidad de los resultados (pueden llegar a conclusiones diferentes con los mismos datos iniciales) ni la autoexplicación de los mismos.

En la cultura popular, se ha extendido el concepto de IA generativa como una caja negra cuyos resultados no son garantía de nada (Ocampo y Santa Catarina, 2018). Esto es una media verdad. Las IA LLM, entendidas tales como redes neuronales profundas, están formadas por unos elementos enormemente simples, las neuronas artificiales, pero cuando se agrupan en número suficiente, como ocurre en estas grandes redes neuronales, con millones o miles de millones de parámetros, son capaces de generar una complejidad emergente (Ruiz, 2001) que, por definición (manejo de respuesta, entrenamiento, profundidad de la red, etc.) no permite que **a)** den siempre la misma respuesta a idénticas entradas, y **b)** expliquen sin rasgo de alucinación las decisiones que han llevado a los resultados entregados.

Una alucinación ocurre cuando una red neuronal no ha sido bien entrenada en un asunto determinado, y cuando procesa lo que se le solicita, al no tener información suficiente, usa su propia naturaleza estocástica para dar una respuesta, que dentro de su modelo interno puede ser adecuada, pero generando un resultado falso (Huidobro, 2024); es algo que se puede evitar mejorando la calidad de los datos de entrenamiento de la red. Es importante señalar aquí que las redes neuronales no manejan palabras ni las comprenden, sino que gestionan datos de entrenamiento con códigos que traducen las palabras a un lenguaje interno. No hay pensamiento, pero sí resultados acordes al pensamiento, en una red neuronal.

Es posible, con todo, implementar una estructura de redes de aprendizaje profundo que puedan dar resultados fiables en una sala judicial con el debido entrenamiento, y justificarlos; se llaman *redes neuronales supervisadas*. Estas son un modelo de aprendizaje automático que se entrena con datos previamente etiquetados, donde cada ejemplo de entrada tiene asociada su salida correcta (Mariño y Primorac, 2016). Durante el proceso de entrenamiento/aprendizaje profundo, el modelo aprende a hacer predicciones analizando estos ejemplos y ajustando sus parámetros internos para minimizar el error entre sus predicciones y las etiquetas reales. Este tipo de redes se utiliza en diversos campos como la clasificación de imágenes, la predicción de precios y diagnósticos, siendo especialmente efectivas cuando se dispone de conjuntos de datos grandes y bien etiquetados, aunque pueden enfren-

tar desafíos como el sobreajuste y la necesidad de datos de alta calidad. El modelo guiado por datos etiquetados da a la red una referencia clara de lo que se considera *correcto*. Durante el entrenamiento, se ajusta el proceso para producir salidas lo más cercanas posible a las respuestas correctas. Esto limita la posibilidad de generar respuestas erróneas o alucinaciones, ya que el modelo es penalizado si predice algo incorrecto. Gracias al uso de los datos bien etiquetados, los errores se reducen hasta ser casi despreciables, ya que se está aprendiendo de ejemplos verídicos, en lugar de intentar encontrar pautas en estructuras en datos sin ningún tipo de guía explícita, caso de las redes no supervisadas. De nuevo, el positivismo jurídico podría argumentar la posibilidad de unos resultados demasiado inflexibles, probablemente con razón.

La primera pregunta, por tanto, que hemos de responder es para qué podríamos usar una IA basada en una red supervisada en una sala de juicios. ¿Podríamos comprobar su fiabilidad? Precisamente abundaremos en ello en la discusión de este artículo. Una de las mejores fuentes de datos de entrenamiento legislativo bien etiquetado y de gran volumen es la legislación civil española. Precisamente, la metodología aplicada, basada en la consulta de literatura complementaria y especializada, nos lleva a poder plantear soluciones viables hacia una propuesta de trabajo que podría llevar a interesantes resultados.

La reciente Ley Orgánica 1/2025, de medidas para la Eficiencia del Servicio Público de Justicia, introduce cambios sustantivos en la organización judicial española, eliminando la denominación «Juzgados de Primera Instancia» y unificando competencias bajo la figura de Juzgados Civiles Generales. Este ajuste terminológico responde a un objetivo de simplificación estructural y optimización de recursos, alineado con la digitalización progresiva del sistema (Exposición de Motivos, LO 1/2025). En consecuencia, las referencias del estudio a «Juzgados de Primera Instancia» deben entenderse en el contexto de la nueva configuración legal, donde estos órganos asumen funciones equivalentes, pero bajo una nomenclatura actualizada. Este cambio no altera la idoneidad experimental de la jurisdicción civil, pues la homogeneidad procedimental y la dependencia de pruebas documentales persisten como rasgos definitorios. Sin embargo, obliga a recalibrar el alcance geográfico y competencial de la propuesta, asegurando que los modelos de IA se adapten a las directrices de eficiencia y especialización previstas en la normativa vigente.

2. Discusión

En este artículo presentamos una propuesta experimental que podría ser fácilmente implementable, y que permitiría llegar a conclusiones interesantes sobre las que trabajar desde el punto de vista del positivismo jurídico. Se ha considerado para tal propuesta la jurisdicción civil (LOPJ, 1985), ya que con ella se podrían generar datos de entrenamiento robustos que puedan ser etiquetados, con ejemplos de entrada-salida (la legislación civil es en ese sentido un excelente banco de pruebas, dada su amplitud y extenso corpus).

Parece más sencillo y conveniente implementar experimentalmente una red neuronal supervisada en los Juzgados de Primera Instancia, debido a la naturaleza estructurada y extensa de la legislación que los regula. En este ámbito, el Código Civil Español (y leyes especiales que lo complementan) y la Ley de Enjuiciamiento Civil tienen el potencial de actuar como una base de datos centralizada y jerárquica, que organizan de manera clara y precisa las normas aplicables a cada situación. Esto facilita que un modelo supervisado pueda identificar pautas y patrones claros entre los hechos presentados en un caso y los artículos legales relevantes al asunto.

Los procedimientos civiles suelen presentar asimismo una estructura más homogénea, objetiva y predecible en comparación con otros tipos de litigios. Las decisiones se basan principalmente en pruebas documentales (como contratos, testamentos o facturas) y en la aplicación directa de normas legales específicas. Esto reduce la influencia de factores subjetivos, como las intenciones emocionales o abstractas, que suelen ser relevantes en procesos penales o de familia. Poniendo un ejemplo, en un caso de incumplimiento de contrato, el tribunal evalúa principalmente si las cláusulas se cumplieron o no, lo que simplifica el razonamiento legal y lo hace más adecuado para un modelo supervisado que busca relaciones claras y patrones consistentes.

La caracterización de los procedimientos civiles como procesos homogéneos, objetivos y predecibles, basados en pruebas documentales y aplicación directa de normas legales, se alinea con postulados fundamentales de la doctrina procesalista. Autores como José María Asencio Mellado (2018) destacan que el derecho procesal civil, a diferencia de otras jurisdicciones, se estructura en torno a principios de escrituración y formalismo probatorio, don-

de la valoración de documentos (contratos, facturas, testamentos) ocupa un rol central, reduciendo la subjetividad inherente a la interpretación de intenciones o contextos emocionales. Esta perspectiva es reforzada por Eduardo Couture (1958), quien enfatiza que la predictibilidad en el ámbito civil deriva de la naturaleza dispositiva del proceso, donde las partes delimitan el conflicto mediante alegaciones y pruebas concretas, facilitando una aplicación mecánica de las normas. En España, esta visión se consolida con la Ley de Enjuiciamiento Civil (LEC 1/2000), que en su artículo 265 establece la primacía de la prueba documental sobre otras formas de prueba en casos patrimoniales, legitimando así un marco procedimental estandarizado. Tales fundamentos doctrinales y normativos respaldan la idoneidad de los juzgados civiles como campo experimental para modelos de IA supervisada, ya que la estructura formalizada y la dependencia de patrones claros (e. g., cumplimiento de cláusulas contractuales) permiten entrenar algoritmos con datos etiquetados coherentes, minimizando ambigüedades. Ignorar estas bases teóricas supondría omitir el sustrato jurídico que justifica la elección metodológica, razón por la cual se integran aquí como sustento crítico de la propuesta.

Otro factor importante es la amplia disponibilidad de datos históricos etiquetados. En los juzgados civiles, las resoluciones suelen estar bien documentadas y publicadas en bases de datos como el Centro de Documentación Judicial, conocido por el acrónimo CENDOJ, creado por el Consejo General del Poder Judicial (CGPJ, 1977). Estas resoluciones incluyen tanto los argumentos de las partes como el resultado final, proporcionando entradas (hechos y pruebas) y salidas (decisiones y normativas aplicadas) bien definidas.

Reforzaría, asimismo, la conveniencia de probar un modelo supervisado en este ámbito: la consistencia de las resoluciones judiciales. Los jueces civiles suelen basar sus decisiones en precedentes claros y en la aplicación directa de la jurisprudencia. Ello implica que casos con características similares suelen resolverse de manera análoga, lo que permitiría al modelo identificar patrones recurrentes y realizar predicciones basadas en datos históricos. Los procedimientos civiles, por otro lado, presentan un menor impacto ético y social en caso de errores. Mientras que en derecho penal un error podría implicar la privación de libertad de una persona, en el ámbito civil los errores suelen estar relacionados con disputas económicas o contractuales. Ello convierte a los Juzgados de Primera Ins-

tancia en un campo de pruebas más seguro desde todos los puntos de vista para evaluar la eficacia y confiabilidad de una red neuronal supervisada.

Nieva Fenoll ha abordado críticamente la tensión entre automatización y garantías procesales, subrayando que la IA no debe entenderse como un sustituto de la interpretación judicial, sino como una herramienta auxiliar sujeta a principios como la contradicción y la tutela judicial efectiva (Nieva Fenoll, 2021). Desde una perspectiva similar, trabajos como los de Montero Aroca (2019) analizan cómo la predictibilidad algorítmica puede coexistir con la flexibilidad requerida en el derecho procesal civil, siempre que se preserve la capacidad de los jueces para modular decisiones en casos excepcionales. Estudios recientes han explorado específicamente la posibilidad del uso de la IA en procesos monitorios civiles, destacando su potencial para agilizar la gestión de reclamaciones masivas (Esteve, 2024). Estos enfoques refuerzan la necesidad de un marco ético-normativo que, como sugiere la presente propuesta, priorice la transparencia y el control humano. La omisión inicial de estas referencias se subsana aquí, integrando un diálogo interdisciplinar que fortalece la solidez teórica del trabajo.

Esta propuesta, es importante indicarlo, busca la realización de trabajos experimentales que no interfieran ni influyan en los juzgados donde se apliquen, para ir obteniendo datos de cotejo (decisión juez vs. decisión de IA) que puedan usarse para analizar resultados y eficacia. Se trata solo de un proceso experimental de toma de datos, que serían usados para estudios científicos hacia la mejora del contraste del uso de las IA en una eventual asistencia judicial a la toma de decisiones que aumente la consistencia en la aplicación de la ley. Debidamente preentrenado un modelo de IA, debería dar una respuesta que se compararía con la sentencia del tribunal, llegando a conclusiones que se almacenarían en una base de datos, que al final del proceso experimental (probablemente unos 12 meses) supondría el suficiente *big data* como para ir adelantando posibles conclusiones y estrategias. Repitiendo el experimento en diversas salas, de forma transparente y confidencial, se podrían obtener unos resultados válidos para realizar estudios de eficiencia y sobre todo buscando una respuesta heurística a la gran duda de la fiabilidad de estas tecnologías, superando el *hándicap* de la caja negra.

El objetivo principal de tales pruebas se centraría en identificar las fortalezas y debilidades de las IA en la

práctica jurídica, proporcionando una base empírica para futuras mejoras y ajustes y una retroalimentación hacia los modelos, de cara a su mejora y ulterior entrenamiento optimizado. Al comparar las actuaciones de los profesionales humanos con las respuestas generadas por las IA, se podrá determinar en qué medida la tecnología actual en su *estado del arte* puede complementar y apoyar el trabajo de los jueces. Se podría ir testando el modelo en casos relativamente sencillos para poco a poco ir aumentando la complejidad.

¿Qué tipo de tecnología sería la más interesante para implementar una red neuronal supervisada en una sala de Primera Instancia? La técnica transformadora (Vaswani *et al.*, 2017), en la que se basan modelos extensos de lenguaje como ChatGPT, ha demostrado de sobras su robustez y su idoneidad para el manejo del lenguaje jurídico de la legislación civil, por lo que parece próxima a los objetivos planteados por su capacidad de manejo del lenguaje. Actualmente, transformadores como BERT, RoBERTa (Adoma *et al.*, 2020) o, incluso, modelos preentrenados en dominios legales (como LegalBERT, descrito en Mamakas *et al.*, 2022) son ideales para comprender y analizar documentos legales, identificar conceptos clave y establecer relaciones entre ellos, lo que los pone muy cerca de la implementación buscada: la legislación civil, o un subconjunto de ella.

Por otro lado, los transformadores pueden manejar información contextual más profunda que las redes tradicionales, lo que permite evaluar los hechos ocurridos, las leyes civiles aplicables y argumentos presentados en los documentos de las partes, pero, sobre todo, poder interrogar a la propia red sobre las decisiones finales. Los modelos supervisados pueden ser entrenados para predecir decisiones basándose en casos previos. Por ejemplo, clasificar si un caso debe resolverse a favor del demandante o demandado, o incluso sugerir indemnizaciones basándose en patrones históricos. Con el paso del tiempo, tras este proceso heurístico, la tecnología probada podría funcionar como una herramienta de análisis para agilizar el trabajo del juez de primera instancia, garantizar mayor consistencia y encontrar patrones en la jurisprudencia.

Los modelos transformadores, además, son ideales para ofrecer una explicación del proceso decisorio, debido a su mecanismo de atención característico. Modelos como BERT ofrecen mecanismos de atención que indican qué

partes del texto consideran más relevantes para su decisión. Usando estos pesos de atención, es posible identificar las frases o palabras clave que el modelo utiliza para llegar a una conclusión.

Por ejemplo, en un caso de conflicto por un contrato, si el modelo decide que el contrato es vinculante según el artículo 1255 del Código Civil, el informe podría resaltar las partes del texto donde el modelo encontró términos como «vinculante», «contrato», citas de jurisprudencia relevante, o términos conflictivos que le han llevado a tal decisión. En el siguiente párrafo vemos una posible respuesta razonada obtenida por los autores con ChatGPT 4.0.

Ejemplo de razonamiento:

- *El incumplimiento de contrato por parte del demandado no queda amparado bajo el artículo 1105 del Código Civil, ya que la imposibilidad de cumplimiento no fue imprevisible ni inevitable.*
- *Según el artículo 1124, el demandante tiene derecho a resolver el contrato y exigir daños y perjuicios.*
- *Casos similares como la Sentencia TS ###/#### avallan esta interpretación.*

Naturalmente, se trata de un caso hipotético. ChatGPT es un modelo de amplio espectro, pero limitado en el caso que nos ocupa.

También es posible añadir una etapa en la que el modelo genere un resumen del caso, destacando los argumentos del demandante y el demandado, los artículos de ley relevantes y cómo estos se aplican al caso en cuestión. Para esta tarea, se puede usar un transformador especializado en generación de texto, como T5 o GPT, que explique el razonamiento en lenguaje natural (Zaki, 2024).

Cabría asimismo una explicabilidad *post hoc* (González, 2024) del proceso de razonamiento seguido, mediante herramientas externas a la red supervisada, en una suerte de análisis forense de decisiones, ideal para un proceso experimental preliminar como el indicado en este texto. Herramientas como LIME (Local Interpretable Model-Agnostic Explanations) o SHAP (SHapley Additive exPlanations) se pueden usar para explicar decisiones de modelos de aprendizaje profundo en lenguaje humano (Assegie, 2023). Todo ello podría ser de utilidad para la elaboración de estrategias de trabajo y documentación para la mejora de los sistemas, con el fin de poder dar un *feedback* al juez que combinara una sugerencia de veredicto y un razonamiento de cómo ha llegado la red neuronal a tal decisión. Este proceso podría permitir al juez comparar su universo decisorio con el que plantearía la red neuronal y poder actuar en consecuencia. En la figura 1, se muestra un grafo con el flujo de trabajo posible que implicaría esta primera iniciativa de implementación experimental de una red transformadora supervisada en Juzgados de Primera Instancia. El código para generar la gráfica se añade como anexo al final de este trabajo.

Figura 1. Flujo de una red supervisada transformadora en Juzgados de Primera Instancia



Fuente: elaboración propia

El flujo representado en la gráfica sigue un camino lineal entre los nodos indicados, de derecha a izquierda y de abajo arriba:

1. Recopila el texto del caso. Este es el punto inicial del flujo. Se introduce el texto completo del caso, que incluye detalles como la demanda del demandante, los argumentos de las partes, los hechos presentados y cualquier otra documentación relevante. Siguiendo el

ejemplo utilizado, podría ser un texto que describiera una disputa sobre un contrato incumplido.

2. Preprocesa los datos para adaptarlos al modelo. Antes de que el modelo pueda trabajar con el texto, se realiza un preprocesamiento automatizado. Esto incluye tareas como: tokenización (dividir el texto en palabras, frases o tokens más pequeños para que el modelo los interprete); limpieza del texto (eliminación de elemen-

tos irrelevantes como caracteres especiales o datos innecesarios para facilitar el análisis) y conversión en vectores (traducción del texto en números que el modelo pueda procesar; este paso deviene fundamental para que los datos estén en un formato adecuado para el transformador).

3. El modelo analiza el caso y hace una predicción supervisada. Este modelo, entrenado específicamente en textos legales (como LegalBERT), analiza la entrada para identificar patrones y correlaciones en el texto del caso. Busca los hechos más relevantes, los argumentos principales y las normativas legales que podrían aplicarse al caso. Es el componente que lleva a cabo la «predicción supervisada» que define este modelo.
4. Basándose en el entrenamiento previo, el modelo genera una predicción preliminar. Esto podría ser el resultado esperado del caso, como: «El demandado debe indemnizar al demandante con 10.000 euros por incumplimiento contractual». Este paso equivaldría a una «propuesta» del modelo para resolver el caso, basada en datos históricos y patrones aprendidos.
5. Se genera un informe que explica el razonamiento detrás de la predicción. Una vez que el modelo ha hecho su predicción, se activa un componente adicional para generar un informe razonado. Este informe explica qué artículos legales o jurisprudencia respaldan la decisión, qué argumentos fueron clave en el razonamiento y cómo se aplicaron las normas al caso específico.
6. La salida incluye tanto el veredicto preliminar (o predicción del modelo) como el razonamiento detallado, que el juez puede revisar y utilizar. Este paso le da al juez una base clara y fundamentada para tomar una decisión definitiva. El juez, naturalmente, puede aceptar, modificar o rechazar la propuesta del modelo, utilizando el informe como una herramienta de apoyo en su análisis.

En condiciones ideales, este podría ser el funcionamiento de un modelo piloto de IA aplicada mediante redes neuronales profundas tipo transformador supervisadas en Juzgados de Primera Instancia, de cara a obtener datos heurísticos que puedan dar una respuesta a su fiabilidad o su posible mejora. En el Anexo II de este artículo se muestra una sugerencia de posibles campos de utilidad a la hora de implementar este experimento en una sala judicial.

Los autores quisieran enfatizar, como cierre de esta discusión, que la distinción entre el contexto de descubrimiento y el contexto de justificación (Reichenbach, 1938) resulta fundamental para clarificar los alcances y limitaciones de la propuesta presentada. En primer lugar, el análisis de jurisprudencia y legislación mediante redes neuronales supervisadas se enmarca predominantemente en el contexto de descubrimiento, ya que el modelo identifica patrones, correlaciones y tendencias en datos históricos para predecir resoluciones basadas en casos similares. Este enfoque se alinea con la naturaleza inductiva de las tecnologías de big data, que buscan generalizar a partir de ejemplos concretos, detectando regularidades estadísticas. Por ejemplo, al entrenar la IA con sentencias etiquetadas del CENDOJ, el sistema aprende a asociar hechos específicos (como incumplimientos contractuales) con resultados judiciales típicos (indemnizaciones o resoluciones de contrato), lo que permite anticipar decisiones futuras. Sin embargo, como señala el revisor, este proceso no garantiza per se la corrección jurídica, pues se centra en la frecuencia y coherencia histórica, no en la validación normativa.

No obstante, la propuesta integra elementos del contexto de justificación mediante dos mecanismos clave. Por un lado, la supervisión basada en datos etiquetados -donde las «respuestas correctas» corresponden a resoluciones judiciales ya validadas- introduce criterios externos de corrección, acercando el modelo a sistemas expertos tradicionales. Aquí, la red neuronal no solo descubre patrones, sino que se ajusta a un marco normativo predefinido (el Código Civil, la LEC, etc.), evaluando sus predicciones frente a estándares jurídicos explícitos. Por otro lado, herramientas como los mecanismos de atención de los transformadores (que destacan artículos legales relevantes) y métodos *post hoc* (LIME, SHAP) permiten traducir las decisiones de la IA en razonamientos estructurados, vinculando las predicciones a normas concretas y jurisprudencia. Esto trasciende la mera detección de regularidades, al ofrecer una justificación basada en fuentes del derecho, lo que acerca la tecnología al ámbito de la validación procedimental. Así, mientras el entrenamiento y la predicción operan en el contexto de descubrimiento, la explicabilidad y el contraste con decisiones humanas introducen una dimensión justificativa, asegurando que las salidas de la IA no sean meramente estadísticas, sino jurídicamente fundamentadas.

La propuesta aquí presentada, por tanto, oscila deliberadamente entre ambos contextos: utiliza el descubrimiento

para agilizar el análisis de casos y predecir resultados, pero incorpora la justificación para garantizar que dichas predicciones se alineen con el marco legal vigente. Esta dualidad refleja la naturaleza híbrida de la IA supervisada en el ámbito judicial, donde la eficiencia predictiva debe coexistir con la rigurosidad normativa. Aclarar esta interacción no solo enriquece el marco teórico del trabajo, sino que subraya cómo la tecnología puede servir como puente entre la práctica inductiva y la validación deductiva, superando la dicotomía entre ambos contextos.

Conclusión

Los modelos de inteligencia artificial supervisados representan una herramienta con enormes potencialidades para el orden judicial civil, no como un reemplazo de los jueces, sino como un complemento que puede agilizar procesos, aumentar la consistencia en las resoluciones y ofre-

cer análisis más sistemáticos de los casos. La jurisdicción civil se perfila como el terreno ideal para experimentar esta implementación tecnológica debido a su estructura normativa clara, la objetividad de sus procesos y la menor complejidad ética de sus resoluciones (que no se niega). Los modelos transformadores, especialmente aquellos entrenados en dominios legales, emergen como la tecnología más prometedora, capaz no sólo de analizar documentos legales con precisión, sino también de explicar sus razonamientos mediante mecanismos de atención o post hoc con herramientas como LIME o SHAP. Esta propuesta experimental podría plantearse con el debido presupuesto y plazos en varios juzgados civiles, para obtener resultados que ayudaran a mejorar los modelos y a establecer protocolos rigurosos de implementación, así como de supervisión y análisis desde la filosofía del derecho, que coadyuvaran a garantizar la fiabilidad y transparencia de estas tecnologías emergentes.

Referencias bibliográficas

- ADOMA, A. F.; HENRY, N. M.; CHEN, W. (2020). «Comparative analyses of bert, roberta, distilbert, and xlnet for text-based emotion recognition». *IEE*, págs. 117-121. 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP).
- ASENCIO MELLADO, J. M. (2018). *Derecho procesal civil: Doctrina y aplicaciones*. Tirant lo Blanch.
- CORES, O. B. (2023). «Repercusiones de la Inteligencia Artificial en la legitimidad del sistema penal: Una reflexión desde la justicia procedimental». *Indret: Revista para el Análisis del Derecho*, Vol. 4, n.º 12.
- COUTURE, E. (1958). *Fundamentos del derecho procesal civil*. Depalma.
- ESTEVE, V. C. (2024). «La afectación de la seguridad jurídica del consumidor en los contratos de crédito online y en las reclamaciones mediante el proceso monitorio». *Boletín del Ministerio de Justicia*, vol. 78, n.º 2275, págs. 1-39.
- Ley 1/2000, de 7 de enero, de Enjuiciamiento Civil, art. 265. (2000). BOE núm. 7, de 8 de enero de 2000 [en línea]. Disponible en: <https://www.boe.es/buscar/act.php?id=BOE-A-2000-323>
- Ley Orgánica 1/2025, de 10 de marzo, de medidas para la Eficiencia del Servicio Público de Justicia. (2025). BOE núm. 62, de 11 de marzo de 2025. [en línea]. Disponible en: <https://www.boe.es/buscar/act.php?id=BOE-A-2025-76>
- LOPJ (1985). Ley Orgánica 6/1985, de 1 de julio, del Poder Judicial. Boletín Oficial del Estado. [en línea]. Disponible en: <http://tinyurl.com/mr39b5tm>
- Consejo General del Poder Judicial, CGPJ. (1997). Acuerdo de 7 de mayo de 1997, del Pleno del Consejo General del Poder Judicial, por el que se aprueba el Reglamento 1/1997, del Centro de Documentación Judicial. [en línea]. Disponible en <https://www.boe.es/buscar/doc.php?id=BOE-A-1997-11001>.
- NIEVA FENOLL, J. (2021). *Inteligencia artificial y proceso judicial: ¿Hacia un modelo de justicia predictiva?* Aranzadi.
- NIEVA FENOLL, J. (2022). «Un cambio generacional en el proceso judicial: la inteligencia artificial». y *justicia digital*, n.º 153.
- GONZÁLEZ, A. S. (2024). *Validación de explicabilidad post-hoc de modelos de tipo caja-negra usando grandes modelos de lenguaje*. Tesis. E.T.S. de Ingenieros Informáticos (UPM)
- HUIDOBRO, C. B. (2024). «Alucinaciones de la inteligencia artificial: impacto en tecnología, política y sociedad». *Revista Estrategia, Poder y Desarrollo*, vol. 3, n.º 5, págs. 47-64. DOI: <https://doi.org/10.25062/2955-0289.4847>
- MAMAKAS, D.; Tsotsi, P.; Androutsopoulos, I.; Chalkidis, I. (2022). «Processing long legal documents with pre-trained transformers: Modding legalbert and longformer». *arXiv*. DOI: <https://doi.org/10.18653/v1/2022.nllp-1.11>
- MARIÑO, S. I.; PRIMORAC, C. R. (2016). «Propuesta metodológica para desarrollo de modelos de redes neuronales artificiales supervisadas». *IJERI: International Journal of Educational Research and Innovation*, vol. 6, págs. 231-245.
- MONTERO AROCA, J. (2019). *Derecho jurisdiccional III: Proceso civil* (25ª ed.). Tirant lo Blanch.
- MORESO, J. J. (2017). «Positivismo jurídico y filosofía analítica. Teoría & Derecho». *Revista de pensamiento jurídico*, vol. 22, págs. 119-136.
- OCAMPO, M.; SANTA CATARINA, C. (2018). «Inteligencia artificial». Oficina de información científica y tecnologica para el congreso de la unión, págs. 1-2.

- PRAKKEN, H.; SARTOR, G. (2004). «The three faces of defeasibility in the law». *Ratio Juris*, vol. 17, n.º1, págs. 118-139. DOI: <https://doi.org/10.1111/j.0952-1917.2004.00259.x>
- REICHENBACH, H. (1938). «Experience and prediction: An analysis of the foundations and the structure of knowledge». *Phoenix Books*. The University of Chicago Press.
- RUIZ, D. F. (2001). «Debate-Inteligencia humana e inteligencia artificial». *Quark*, pág. 109-112.
- TORRADO, J. M. S. (2022). «Inteligencia Artificial y Filosofía del Derecho». *E-Revista Internacional de la Protección Social*, vol. 7, n.º 212.
- VASWANI, A. et al., (2017). «Attention is all you need». *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, págs. 6000-6010 [en línea]. Disponible en: <https://dl.acm.org/doi/10.5555/3295222.3295349>
- XU, Y.; LIU, X.; CAO, X.; HUANG, C.; LIU, E.; QIAN, S.; ZHANG, J., et al. (2021). «Artificial intelligence: A powerful paradigm for scientific research». *The Innovation*, vol. 2, n.º 4. DOI: <https://doi.org/10.1016/j.xinn.2021.100179>
- ZAKI, M. Z. (2024). «Revolutionising translation technology: a comparative study of variant transformer models - BERT, GPT AND T5». *Computer Science and Engineering-An International Journal*, vol. 14, n.º 3, págs. 15-27. DOI: <https://doi.org/10.5121/cseij.2024.14302>

Anexo I

Código Python para representación del grafo de figura 1.

```
import matplotlib.pyplot as plt
import networkx as nx

# Crear un grafo para representar los componentes de una red supervisada
# aplicada en salas civiles
G = nx.DiGraph()

# Nodos principales
nodes = [
    "Entrada: Texto del caso\n(demanda, argumentos)",
    "Preprocesamiento\n(tokenización, limpieza)",
    "Modelo Transformer\n(BERT, LegalBERT)",
    "Predicción del modelo\n(veredicto preliminar)",
    "Generación de informe\n(jurisprudencia, razonamiento)",
    "Salida para el juez\n(veredicto + razonamiento)"]

# Añadir nodos al grafo
for node in nodes:
    G.add_node(node)

# Añadir relaciones entre los nodos
edges = [
    ("Entrada: Texto del caso\n(demanda, argumentos)", "Preprocesamiento\n(-
tokenización, limpieza)"),
    ("Preprocesamiento\n(tokenización, limpieza)", "Modelo Transformer\n(-
BERT, LegalBERT)"),
    ("Modelo Transformer\n(BERT, LegalBERT)", "Predicción del modelo\n(ver-
edicto preliminar)"),
    ("Predicción del modelo\n(veredicto preliminar)", "Generación de infor-
me\n(jurisprudencia, razonamiento)"),
    ("Generación de informe\n(jurisprudencia, razonamiento)", "Salida para el
juez\n(veredicto + razonamiento)")]

for edge in edges:
    G.add_edge(*edge)

# Configurar el layout del grafo para un flujo de izquierda a derecha
pos = nx.spring_layout(G, seed=42, k=1, iterations=50)

# Configuración de la figura
plt.figure(figsize=(16, 10)) # Aumentar el tamaño de la figura

# Dibujar los nodos y las aristas
nx.draw_networkx_nodes(G, pos, node_size=6000, node_color='skyblue', edgeco-
lors='black', node_shape='o', alpha=0.8)
```

```

nx.draw_networkx_edges(G, pos, arrowstyle="->", arrowsize=20, edge_color='gray', width=2)

# Dibujar etiquetas con fuente más grande y ajustada
nx.draw_networkx_labels(G, pos, font_size=12, font_weight='bold', font_color='black', verticalalignment='center', horizontalalignment='center')

# Título del gráfico
plt.title("Flujo de una Red Supervisada Transformer en Juzgados de Primera Instancia", fontsize=16, fontweight='bold')

plt.axis("off")

# Mostrar la gráfica
plt.tight_layout()

plt.show()

```

Anexo II

Para almacenar datos en una base de datos orientada a un experimento heurístico que contraste las decisiones tomadas por una IA y las de un juez humano en procedimientos civiles, los autores proponen la siguiente información estructurada:

Metadatos del caso:

- ID del caso: identificador único del caso.
- Información del tribunal: nombre y ubicación del tribunal donde se llevó el caso.
- Fecha del fallo: fecha en que se emitió la decisión.
- Detalles del juez: identificador del juez humano que presidió el caso (con fines de anonimización; no serán usados, pero pueden ser de utilidad informativa para el propio juez a la hora de dar *feedback*, como se verá en el penúltimo punto de este anexo II).

Entradas para el modelo:

- Texto del caso: descripción completa del caso, que incluya:
- Demandas del demandante.
- Argumentos y contraargumentos de las partes involucradas.
- Pruebas presentadas (por ejemplo, contratos, facturas u otros documentos relevantes).
- Referencias legales: artículos de la ley citados por las partes o relevantes para el caso.
- Metadatos contextuales: factores como la jurisdicción y el tipo de caso (e. g., incumplimiento de contrato, disputa de herencia).

Salidas de la IA:

- Decisión predicha: el fallo propuesto por la IA (e. g., qué parte debería ganar, daños asignados, etc.).
- Explicación: razones clave de la decisión de la IA, incluyendo:
- Artículos o jurisprudencia aplicados.
- Frases o textos destacados considerados críticos (utilizando mecanismos de atención del modelo).
- Nivel de confianza: grado de confianza de la IA en su decisión.

Salidas del juez:

- Decisión del juez: el fallo real emitido por el juez humano.
- Razonamiento: explicación legal y fáctica proporcionada por el juez para justificar su decisión (en lenguaje natural, extracción directa del documento del fallo).
- Referencias utilizadas: artículos de ley, precedentes u otros documentos citados en el fallo.

Datos de contraste:

- Coincidencia/desacuerdo: métrica binaria o porcentual que indique si las decisiones de la IA y del juez coinciden y si es necesario el grado de coincidencia, usando una métrica por determinar.
- Diferencias clave: discrepancias específicas en el razonamiento o en los resultados entre la IA y el juez humano.
- Análisis de errores: casos de «alucinaciones» de la IA o desviaciones significativas de las normas legales.

Metadatos del experimento:

- Información del modelo: detalles sobre la IA utilizada (e. g., versión, características de los datos de entrenamiento y arquitectura usada como transformadora o LegalBERT, entre las ya comentadas).
- Referencias de entrenamiento: conjunto de datos o corpus utilizado para entrenar la IA.
- Métricas de evaluación: criterios utilizados para medir la alineación (e. g., precisión, exhaustividad, explicabilidad).

Registro y retroalimentación:

- Registros de sesión: detalles del preprocesamiento, tokenización y pasos intermedios realizados por la IA.
- *Feedback* de los jueces: evaluación cualitativa del juez sobre el razonamiento de la IA, sugerencias de mejora o áreas de preocupación. Este punto es vital, pues el juez puede valorar desde su área de conocimiento los posibles errores o sesgos que ha visto en la IA.

Seguridad y privacidad:

- Indicadores de anonimización: información sobre si los datos sensibles del caso han sido anonimizados.
- Permisos de uso de datos: autorizaciones para usar los datos en investigaciones o en futuros entrenamientos de IA.

Cita recomendada

CAJA MOYA, Cristina; QUIROGA RODRÍGUEZ, Elio (2025). «Más allá de la caja negra: la IA en los juzgados del orden civil (una propuesta heurística)». *IDP. Revista de Internet, Derecho y Política*, núm. 43. UOC. [Fecha de consulta: dd/mm/aa]. DOI: <http://dx.doi.org/10.7238/idp.v0i43.432360>



Los textos publicados en esta revista están –si no se indica lo contrario– bajo una licencia Reconocimiento-Sin obras derivadas 3.0 España de Creative Commons. Puede copiarlos, distribuirlos y comunicarlos públicamente siempre que cite su autor y la revista y la institución que los publica (*IDP. Revista de Internet, Derecho y Política*; UOC); no haga con ellos obras derivadas. La licencia completa se puede consultar en: <http://creativecommons.org/licenses/by-nd/3.0/es/deed.es>.

Sobre las autorías

Cristina Caja Moya

Universidad del Atlántico Medio. Facultad de Ciencias Sociales y Jurídicas
cristina.caja@pdi.atlanticomedio.es

Es jueza desde 2007 y doctora por la Universidad de Las Palmas de Gran Canaria, con una tesis sobre mediación en accidentes de circulación. Es docente de Derecho Civil en la Universidad del Atlántico Medio desde 2019. Especialista en mediación en conflictos familiares, civiles, mercantiles y de tráfico, así como en extranjería y seguridad vial. Ha sido coorganizadora del Foro de Paz UNESCO-Gran Canaria (2016-2017) y del Congreso Mundial de Arbitraje (2018), además de coordinadora y jurado en el I Legal Hackathon AMETI. Es delegada en Canarias del Foro Internacional de Mediadores Profesionales (FIMEP) y miembro de ORD Latinoamérica. Autora de publicaciones sobre mediación, derecho internacional y derecho civil, incluyendo estudios sobre la relación entre aseguradoras y los objetivos de desarrollo sostenible (ODS). Ha participado en congresos y cursos de especialización en mediación civil, mercantil, sanitaria y en compañías aseguradoras, así como en derecho marítimo y arbitraje. Actualmente es docente e investigadora en la Universidad del Atlántico Medio.

Elio Quiroga Rodríguez

Universidad del Atlántico Medio. Facultad de Ciencias Sociales y Jurídicas
elio.quiroga@pdi.atlanticomedio.es

Ingeniero técnico en Informática de Sistemas por la Universidad de Las Palmas de Gran Canaria, es experto en Astronomía por la Universidad Católica de Murcia, máster en Astronomía y Astrofísica por la Universidad Internacional de Valencia, así como doctorando en Ciencias Sociales por la Universidad Católica de Murcia. Asimismo, Quiroga es guionista, director y productor de cine, así como novelista y autor literario, ganador de premios como el Minotauro de Novela, el Premio de Novela Policía Nacional, Premio de Novela Histórica de Vallirana, Accésit al Premio Everis de Ensayo o finalista al Premio Fernando Lara de Novela. Ha impartido cursos en el CCCB, la Universidad de Las Palmas de Gran Canaria, la Makkers School o el Cabildo de Gran Canaria, y ha dirigido eventos como el Canarias Mediafest. Actualmente es docente e investigador en la Universidad del Atlántico Medio.

