



Jueces robot, dos aproximaciones prácticas a su concepto

Javier Ercilla García

Magistrado del Juzgado de lo Social n.º 10 - Las Palmas de Gran Canaria (España)
j.ercilla@poderjudicial.com | <https://orcid.org/0009-0006-5930-2574>

Extracto

En la era digital actual, la posibilidad de integrar sistemas de inteligencia artificial (IA) en el ámbito jurídico ha suscitado un debate profundo sobre la administración automatizada de la justicia. Este artículo explora dos enfoques paradigmáticos en la conceptualización de un «juez robot»: el basado en reglas y el adaptativo. Mientras que el primero se centra en una codificación explícita de la ley, garantizando predictibilidad y transparencia, inspirado en AlphaZero, el segundo, inspirado en AlphaGo, se adapta continuamente a la jurisprudencia, ofreciendo flexibilidad y capacidad de evolución. A través de un análisis detallado, se discuten las ventajas, limitaciones y aplicaciones potenciales de ambos modelos. Asimismo, se muestran dos ejemplos concretos de juez robot basados en cada uno de los modelos, Python y Tkinter, para la resolución por IA de pleitos relativos a la extinción del contrato de trabajo por voluntad del trabajador fundamentados en retrasos en el abono de los salarios y en materia de revisión de incapacidades permanentes por mejoría.

Palabras clave: juez robot; inteligencia artificial; Administración de justicia; resolución basada en reglas; aproximación autoadaptativa; codificación; predictibilidad; procesamiento del lenguaje natural; GPT-4.

Recibido: 03-11-2023 / Revisado: 10-05-2024 / Aceptado: 14-05-2024 / Publicado (en avance *online*): 14-06-2024

Cómo citar: Ercilla García, J. (2024). Jueces robot, dos aproximaciones prácticas a su concepto. *Revista de Trabajo y Seguridad Social. CEF*, 481, 47-84. <https://doi.org/10.51302/rtss.2024.20111>



Robot Judges, two practical approaches to their concept

Javier Ercilla García

Magistrado del Juzgado de lo Social n.º 10 – Las Palmas de Gran Canaria (España)
j.ercilla@poderjudicial.com | <https://orcid.org/0009-0006-5930-2574>

Abstract

In today's digital era, the possibility of integrating artificial intelligence systems into the legal field has sparked a profound debate about automated justice administration. This article explores two paradigmatic approaches in conceptualizing a "robot judge": the rule-based and the adaptive. While the former focuses on an explicit encoding of the law, ensuring predictability and transparency, inspired by AlphaZero, the latter, inspired by AlphaGo, continuously adapts to jurisprudence, offering flexibility and evolutionary capacity. Through a detailed analysis, the advantages, limitations, and potential applications of both models are discussed. Likewise, two specific examples of a Robot Judge based on each of the models are shown, all based on Python and Tkinter, for AI resolution of lawsuits related to the termination of employment contracts due to delays in salary payments and on the review of permanent disabilities due to improvement.

Keywords: robot judge; artificial intelligence; Justice administration; rule-based resolution; self-adaptive approach; codification; predictability; natural language processing; impact; GPT-4.

Received: 03-11-2023 / Revised: 10-05-2024 / Accepted: 14-05-2024 / Published (online preview): 14-06-2024

Citation: Ercilla García, J. (2024). Robot Judges, two practical approaches to their concept. *Revista de Trabajo y Seguridad Social. CEF*, 481, 47-84. <https://doi.org/10.51302/rtss.2024.20111>

Sumario

1. Introducción
 2. Historia y evolución de AlphaGo y AlphaZero
 - 2.1. AlphaGo: rompiendo barreras
 - 2.2. AlphaZero: hacia la autonomía
 - 2.3. Recapitulación
 3. Concepto de «jueces robot»
 - 3.1. Definición y origen del concepto
 - 3.2. Beneficios potenciales
 - 3.3. Desafíos y consideraciones
 4. Dos ejemplos prácticos de implementación del juez robot
 - 4.1. AlphaJudge
 - 4.1.1. Aplicación
 - 4.1.2. Instrucciones para la inteligencia artificial
 - 4.1.3. Pruebas y evaluación del programa
 - 4.2. AlphaZero Judge
 - 4.2.1. Aplicación
 - 4.2.2. Instrucciones para la inteligencia artificial
 - 4.2.3. Pruebas y evaluación del programa
 5. Estructura y anatomía del programa
 6. Desafíos éticos y técnicos
 - 6.1. Desafíos éticos
 - 6.2. Desafíos técnicos
 7. Conclusiones
 8. Código fuente
- Referencias bibliográficas

1. Introducción

La revolución tecnológica del siglo XXI ha transformado innumerables aspectos de nuestra sociedad, desde la forma en que nos comunicamos hasta la manera en que realizamos negocios. Sin embargo, uno de los dominios que ha permanecido relativamente inalterado por la tecnología es el sistema judicial. A pesar de algunos avances en la digitalización de registros y procedimientos, la esencia de cómo se toman las decisiones judiciales ha permanecido constante durante siglos. Sin embargo, con el advenimiento de la IA avanzada, estamos al borde de una transformación potencialmente radical en este ámbito: la introducción de los jueces robot.

La idea de un juez robot, una entidad que utiliza algoritmos y bases de datos para tomar decisiones judiciales, puede parecer futurista o incluso distópica para algunos. Pero antes de descartar la viabilidad de este concepto, podemos considerar otros dominios en los que la IA ha demostrado capacidades sorprendentes. El juego Go, un antiguo juego de mesa, ha sido durante mucho tiempo considerado un bastión de la capacidad cognitiva humana. Sin embargo, en 2016 el mundo fue testigo de un hito tecnológico cuando AlphaGo, un programa de IA desarrollado por DeepMind, derrotó al campeón mundial, Lee Se-dol, en una serie de partidas (Silver *et al.*, 2016). Este evento no solo demostró el potencial de la IA en tareas complejas, sino que también planteó preguntas sobre cómo podrían aplicarse tecnologías similares en dominios tan cruciales como el judicial. Meses después, AlphaZero derrotaría a AlphaGo.

Lo cierto es que programas como AlphaGo o AlphaZero Go se limitaban a operar dentro de unas reglas muy marcadas: las de un juego de mesa. Sin embargo, la labor judicial, sin perjuicio de regirse por reglas (las leyes), implica una labor cognitiva superior al simple movimiento de fichas en un tablero. Por ende, el concepto de «juez robot» se manifiesta como una evolución natural dada la trayectoria actual de la tecnología y su integración en diversos sectores de la sociedad (Dressel y Farid, 2018). Esta propuesta no solo plantea cuestiones técnicas sobre cómo se podría lograr tal sistema, sino también profundas interrogantes éticas y filosóficas sobre la naturaleza de la justicia y el papel de la humanidad en su administración.

Así, hasta la publicación de GPT-3 en junio de 2020 y sobre todo GPT-4 en marzo de 2023, no teníamos una IA capaz de asumir tareas de cognición profunda en el ámbito del lenguaje. Es decir, la posibilidad de entender el lenguaje, crear lenguaje y actuar en consecuencia. Siendo el lenguaje una de las herramientas esenciales para el juzgador, mediante su comprensión, interpretación y aplicación, el paso hacia los jueces robot se allanaba desde el punto de vista técnico.

La relación entre modelos de IA como AlphaGo y el concepto de «jueces robot» es el foco central de este ensayo, que no podría haberse desarrollado sin la ayuda de grandes modelos del lenguaje (LLM) como GPT-4, Llama-2 o Bard. A través de un análisis exhaustivo, exploraremos la evolución de la IA, desde sus inicios hasta los modelos avanzados de hoy, y cómo estos desarrollos podrían informar o influir en la creación de sistemas judiciales automatizados. Además, abordaremos los desafíos éticos y técnicos asociados con la implementación de la IA en el sistema judicial y reflexionaremos sobre las implicaciones más amplias para la sociedad en su conjunto. Finalmente, se mostrarán dos ejemplos prácticos –desarrollados por quien suscribe– sobre los dos modelos de juez robot que se plantean en el presente artículo.

Es importante destacar que este ensayo no aboga necesariamente por la adopción inmediata de jueces robot, sino por adentrarnos en su estudio y apreciar el estado del arte en relación con la resolución de pleitos.

El sistema judicial es una piedra angular de cualquier sociedad democrática, garantizando la justicia, la equidad y el Estado de derecho. A medida que nos adentramos en una era de tecnología avanzada y en la automatización, es esencial que abordemos cualquier cambio en este dominio con precaución, reflexión y un compromiso inquebrantable con los derechos fundamentales.

2. Historia y evolución de AlphaGo y AlphaZero

El juego Go, originario de China y con más de 2.500 años de antigüedad, ha sido durante mucho tiempo un desafío para la IA debido a su complejidad inherente. A diferencia del ajedrez, donde las computadoras superaron a los humanos desde la victoria de Deep Blue sobre Garry Kasparov en 1997, el Go presentaba un espacio de búsqueda mucho más amplio, lo que lo hace computacionalmente desafiante (Russell y Norvig, 2010). Sin embargo, este panorama cambió drásticamente con la introducción de AlphaGo y, posteriormente, AlphaZero, ambos por DeepMind.

2.1. AlphaGo: rompiendo barreras

Desarrollado por DeepMind, una subsidiaria de Alphabet Inc., AlphaGo representó un avance significativo en el campo de la IA. Utilizando una combinación de redes neuronales y técnicas de búsqueda basadas en árboles Monte Carlo¹, AlphaGo fue entrenado inicialmente en millones de partidas de jugadores humanos, permitiéndole aprender patrones y estrategias comunes en el juego (Silver *et al.*, 2016). Posteriormente, se mejoró mediante el juego contra sí mismo, refinando continuamente sus tácticas y estrategias.

¹ Método para la toma óptima de decisiones en problemas de IA.

El verdadero potencial de AlphaGo se reveló al mundo en 2016, cuando derrotó al campeón mundial Lee Se-dol (9-dan) en una serie de 5 partidas, ganando 4 de ellas (Sang-Hun, 2016). Esta victoria no solo demostró la capacidad de la IA para superar a los humanos en tareas altamente complejas, sino que también destacó el poder de combinar aprendizaje supervisado con aprendizaje por refuerzo (Silver *et al.*, 2016).

2.2. AlphaZero: hacia la autonomía

Si bien AlphaGo representó un avance significativo, DeepMind no se detuvo allí. Introdujeron AlphaZero, una evolución de AlphaGo, con una diferencia clave: en lugar de ser entrenado en partidas humanas, AlphaZero aprendió a jugar Go, ajedrez y shogi simplemente jugando contra sí mismo, sin ningún conocimiento previo de las estrategias humanas (Silver *et al.*, 2017). Esta aproximación de aprendizaje por refuerzo puro permitió a AlphaZero no solo superar a AlphaGo, sino también establecerse como el jugador más fuerte en los 3 juegos mencionados en un tiempo de entrenamiento ciertamente breve.

AlphaZero representa un paso hacia sistemas de IA más autónomos y adaptables. Al no depender de datos humanos, evita posibles sesgos y limitaciones inherentes a esos datos y demuestra una capacidad de automejora que es esencial para aplicaciones más allá de los juegos (Silver *et al.*, 2017).

2.3. Recapitulación

La diferencia esencial por tanto entre AlphaGo y AlphaGo Zero es que la primera aprendió de la experiencia humana previa y se reforzó jugando contra sí misma, con esa base. AlphaGo fue entrenada inicialmente para imitar el juego humano, tratando de igualar los movimientos de los jugadores expertos de juegos históricos registrados, utilizando una base de datos de alrededor de 30 millones de movimientos. Una vez que había alcanzado un cierto grado de habilidad, se entrenó aún más jugando contra otras instancias de sí misma (Schraudolph *et al.*, 1994).

AlphaGo Zero, en cambio, solo aprendió las reglas básicas del juego y ganó experiencia jugando contra sí misma durante 3 días, venciendo a la anterior que había estado entrenando meses. AlphaGo Zero tiene reglas codificadas para establecer hiperparámetros de búsqueda y su red neuronal se actualiza de manera continua.

Lo que se presenta en las líneas superiores es la contraposición entre dos modelos de IA que resuelven los mismos problemas, a saber, la victoria de partidas en juegos de mesa. AlphaGo aprendió mediante el estudio de millones de partidas humanas previas y AlphaZero simplemente aprendió las reglas del juego.

3. Concepto de «jueces robot»

La idea de automatizar procesos y decisiones a través de máquinas no es nueva. Desde la Revolución Industrial, la humanidad ha buscado formas de utilizar la tecnología para mejorar la eficiencia y precisión en diversas áreas. Sin embargo, la propuesta de introducir la automatización en el sistema judicial, específicamente en la figura de un juez robot, es una noción que ha ganado tracción en las últimas décadas, impulsada en gran medida por los avances en IA y aprendizaje automático.

3.1. Definición y origen del concepto

El término «juez robot» se refiere a un sistema automatizado que, utilizando algoritmos y bases de datos, puede tomar decisiones judiciales sin intervención humana directa. Aunque la nomenclatura puede sugerir una entidad física, en la mayoría de los contextos un juez robot es un *software* diseñado para analizar, interpretar y aplicar la ley (Susskind, 2019).

La idea existente detrás de los jueces robot surge de la creciente carga de trabajo en los sistemas judiciales de todo el mundo. Con el aumento de litigios y la complejidad de las leyes modernas, se ha buscado cómo la tecnología podría aliviar parte de esta carga, garantizando al mismo tiempo decisiones procedentes y consistentes (Brownsword, 2008).

3.2. Beneficios potenciales

Uno de los principales argumentos a favor de los jueces robot es la consistencia. A diferencia de los jueces humanos, que pueden ser influenciados por factores externos o sesgos personales, un juez robot tomaría decisiones basadas únicamente en la información proporcionada y en el marco legal establecido. Esto podría llevar a decisiones más consistentes y predecibles, lo que a su vez podría aumentar la confianza en el sistema judicial (Remus y Levy, 2016).

Además, la velocidad y eficiencia son otros beneficios potenciales. Los jueces robot podrían procesar rápidamente grandes cantidades de información, lo que podría reducir significativamente los tiempos de espera en casos judiciales y aliviar la carga de trabajo de los tribunales (Mutlu y Forlizzi, 2008).

3.3. Desafíos y consideraciones

Sin embargo, la idea de delegar decisiones judiciales a máquinas no está exenta de críticas. Uno de los principales desafíos es el riesgo de sesgo. Aunque los jueces robot están

diseñados para ser objetivos, los algoritmos que los alimentan son creados por humanos y pueden, inadvertidamente, introducir sesgos en el sistema. Estos sesgos pueden surgir de los datos utilizados para entrenar al sistema o de las decisiones de diseño tomadas por los programadores (Barocas y Selbst, 2016).

Además, hay preocupaciones sobre la transparencia y la rendición de cuentas. A diferencia de las decisiones humanas, que pueden ser explicadas y justificadas, las decisiones tomadas por algoritmos complejos a menudo son opacas, lo que dificulta entender cómo se llegó a una determinada conclusión (Pasquale, 2015).

Finalmente, hay consideraciones éticas y filosóficas: ¿es éticamente aceptable que una máquina decida sobre asuntos que afectan directamente la vida y libertad de las personas? Además, la ley no es solo un conjunto de reglas, sino también un reflejo de valores y principios sociales. La interpretación y aplicación de la ley requiere un entendimiento profundo del contexto social y cultural, algo que, al menos por ahora, está más allá de las capacidades de cualquier máquina (Moor, 2006; Surden, 2014).

4. Dos ejemplos prácticos de implementación del juez robot

Una vez expuestas las premisas lógicas y técnicas del presente ensayo, vamos a plasmar las mismas en una demostración real y actual de hasta dónde llegaría en la actualidad un juez robot, es decir, cuáles son las actuales capacidades y limitaciones que podríamos encontrar para aplicar, en un sistema judicial como el español, un juez robot. Para ello, se han desarrollado 2 aplicaciones o programas en Python, con la interfaz de Tkinter y con el motor de IA generativa de texto GPT-4.

GPT-4, o Generative Pre-trained Transformer 4, es un modelo de lenguaje autorregresivo basado en la arquitectura *transformer*. Esta arquitectura, introducida por Ashish Vaswani *et al.* (2017) en el artículo «Attention is All You Need», utiliza mecanismos de atención para procesar datos secuenciales de manera paralela, lo que permite un procesamiento más eficiente y una capacidad de modelado más profunda.

GPT-4, al ser una versión más avanzada que GPT-3.5, posee una mayor cantidad de parámetros, lo que le permite tener una capacidad de procesamiento y comprensión del lenguaje más sofisticada. La arquitectura *transformer* en la que se basan ambos modelos permite que, a medida que se incrementa el número de parámetros, el modelo pueda capturar relaciones más complejas y sutiles en los datos. Esto se traduce en una mejor capacidad para generar respuestas coherentes, comprender contextos más amplios y realizar inferencias más precisas.

Comparativamente, GPT-4 puede manejar tareas más complejas y contextos más extensos que GPT-3.5. Además, la capacidad de GPT-4 para generalizar a partir de ejemplos

de entrenamiento es superior, lo que significa que puede adaptarse mejor a tareas para las que no fue específicamente entrenado (Radford *et al.*, 2019; Brown *et al.*, 2020).

Y son esas características de GPT-4, a saber, la comprensión más profunda del lenguaje y la adaptación a tareas no preentrenadas (sin perjuicio de la posibilidad de llevar a cabo un *fine-tuning*), las que permiten a un gran modelo del lenguaje como es GPT-4 llevar a cabo la toma de decisiones judiciales en aquellos supuestos muy específicos que no requieren grandes tareas de cognición profunda.

Así, GPT-4 ha demostrado capacidades notables en una variedad de dominios y tareas, según Bubeck *et al.* (2023), GPT-4 puede resolver tareas novedosas y difíciles que abarcan matemáticas, codificación, visión, medicina, derecho, psicología y más, sin necesidad de instrucciones especiales. Su rendimiento en estas tareas es notablemente cercano al desempeño humano y a menudo supera ampliamente a modelos anteriores como ChatGPT.

Han *et al.* (2023) aplicaron GPT-3.5 y GPT-4 a un problema clásico en el razonamiento inductivo humano conocido como «inducción de propiedades». Aunque GPT-3.5 tuvo dificultades para capturar muchos aspectos del comportamiento humano, GPT-4 fue mucho más exitoso.

Webb *et al.* (2022) realizaron una comparación directa entre razonadores humanos y un modelo de lenguaje grande en una serie de tareas analógicas. Los resultados indicaron que modelos como GPT-3 y GPT-4 han adquirido una capacidad emergente para encontrar soluciones de manera efectiva y sin entrenamiento previo (*Zero-Shot*) a una amplia gama de problemas de analogía.

Almeida *et al.* (2023) investigaron las similitudes y diferencias entre GPT-4 y los humanos en cuanto a razonamiento moral y jurídico. Analizaron aspectos como la atribución de intencionalidad, juicios sobre causación, moralidad de la decepción, fundamentos morales, impacto de la suerte moral en juicios legales, concepto de consentimiento y juicios de violación de reglas. Aunque encontraron altas correlaciones entre las respuestas humanas y de la IA, también identificaron diferencias sistemáticas significativas entre ambas.

Tenemos, por lo tanto, una IA capaz de llevar a cabo tareas complejas de matemáticas, medicina y derecho, una IA capaz de efectuar tareas analógicas y de razonamiento inductivo, es decir, tenemos una IA que puede llevar a cabo someras tareas que debe realizar un juez al dictar una sentencia. Tomando todo ello se desarrollaron 2 aplicaciones para que GPT-4, con una información inicial –unos hechos probados–, fuera capaz de resolver litigios en uno u otro sentido.

Como ya se explicó antes, la premisa de la que parte el presente ensayo es la existencia de 2 modelos de juez robot: el modelo AlphaGo, que aprendió de millones de partidas humanas previas, y el modelo AlphaZero, que aprendió simplemente las reglas y mejoró

jugando contra sí mismo. La asimilación de estos 2 enfoques de resolución de problemas serían el AlphaJudge, que resolvería conforme a la jurisprudencia, esto es, enjuiciamientos humanos previos, y el AlphaZero Judge, que aprendería únicamente la legislación, las reglas y actuaría en consecuencia.

4.1. AlphaJudge

Para este primer enfoque del concepto de «juez robot» hemos tomado un caso concreto de resolución judicial, a saber, el de la extinción del contrato de trabajo por voluntad del trabajador. El artículo 50.1 b) del Estatuto de los Trabajadores dispone que serán causas justas para que el trabajador pueda solicitar la extinción del contrato la falta de pago o retrasos continuados en el abono del salario pactado. Si nos limitáramos a crear un juez robot que resolviera los supuestos de falta de pago, la resolución no precisaría una labor de IA, sino más bien de automatización, dado que la jurisprudencia es bastante clara a la hora de fijar el número de impagos mínimos que harían justa la causa de extinción. Ello, no obstante, la labor de interpretación aún queda para aquellos casos que se sitúan en los límites. Sin embargo, en el ámbito de los retrasos la jurisprudencia es más difusa y requiere llevar a cabo una serie de pasos lógicos y cognitivos que quedarían debidamente integrados en las capacidades de GPT-4 antes señaladas, a saber, la matemática, la analogía y el razonamiento.

La jurisprudencia en el ámbito de la extinción por retrasos en el abono parte de elementos muy dispares y a la par matemáticos, a saber, tiene en cuenta el número de retrasos, el promedio de los mismos, la continuidad, etc.

Consecuentemente, la aplicación desarrollada tendría por objeto facilitar únicamente a la IA la relación de retrasos acreditados, es decir, facilitar únicamente los hechos probados y, una vez con ellos, resolver si procedería o no la extinción del contrato y, por ende, si procedía o no estimar o desestimar la demanda.

AlphaGo, con su enfoque en el aprendizaje supervisado, sería un modelo que toma decisiones basadas en precedentes. Al analizar y aprender de partidas anteriores, AlphaGo desarrolla una comprensión de las estrategias que han resultado exitosas en el pasado. De manera similar, un juez robot basado en dicho modelo podría ser entrenado para analizar y aprender de decisiones judiciales anteriores, desarrollando una comprensión de cómo se han interpretado y aplicado las leyes en casos anteriores (Ashley, 2017).

4.1.1. Aplicación

La aplicación creada parte de un modelo de sentencia, debiendo rellenar los valores no repetitivos, a saber, número de procedimiento, nombres, hechos probados, etc. Siendo la interfaz gráfica la siguiente:

Imagen 1. Imagen global de la aplicación AlphaJudge

Rellenador de modelo

Fecha: 14/9/2023

Número de juicio: []

Nombre del demandante: []

Nombre de la institución demandada: []

¿Comparecen ambas partes?

Sí

No

Hechos probados: []

+ -

Pruebas

Documental Testifical

Pericial Forense

Interrogatorio dte Interrogatorio ddo

Antigüedad: 14/9/2023

Salario día: []

Relación de los retrasos: []

Resolver

Guardar

SENTENCIA

En Las Palmas de Gran Canaria, a [Fecha].

Visto por mí, D. Javier Ercilla García, Magistrado del Juzgado de lo Social n.º 10 de Las Palmas de Gran Canaria, en audiencia pública, el juicio sobre Resolución del contrato de trabajo, seguido ante este Juzgado bajo n.º [Número de juicio], promovido a instancia de [Nombre del demandante], contra [Nombre de la institución demandada] y FOGASA, atendiendo a los siguientes:

ANTECEDENTES DE HECHO

PRIMERO. La parte actora arriba indicada presentó en el decanato una demanda que fue repartida a este Juzgado y en la que, previa alegación de los hechos y fundamentos de derecho que estimó procedentes, terminó suplicando que se dictara una sentencia de conformidad con sus pretensiones.

SEGUNDO. Admitida a trámite la demanda, las partes fueron citadas al acto de juicio. Comparecidas las partes, asistidas en la forma que consta en el acta, se pasó al acto de juicio. En él, y una vez que se hubo efectuado la dación de cuenta de los antecedentes, la parte actora se ratificó en su demanda. La parte demandada se opuso a la demanda y la contestó formulando las alegaciones que constan en acta. Seguidamente, fue abierta la fase probatoria, en la que se practicaron las pruebas que, propuestas por las partes, fueron declaradas pertinentes y constan documentadas en autos. Practicada la prueba, las partes informaron sobre sus pretensiones y el juicio quedó visto para sentencia.

TERCERO. En la sustanciación de estos autos se han observado todas las prescripciones legales aplicables salvo los plazos para dictar sentencia.

HECHO PROBADOS

XXXX. La parte actora no ostenta ni ha ostentado en el año anterior la cualidad de representante unitario o sindical de los trabajadores.

Fuente: elaboración propia a partir de la aplicación.

En la imagen se puede observar una vista previa de la sentencia que se está dictando a la derecha y a la izquierda las distintas entradas, de manera que, conforme se escribe en los recuadros de la izquierda, la sentencia se va actualizando con todos los detalles.

Asimismo, se crea un árbol de decisiones de tal manera que, en aquellas circunstancias en que se pueden plantear varios escenarios, si ocurre uno u otro, el texto de la sentencia cambia. Así ocurre con la opción «¿Comparecen ambas partes?», que pondrá en los antecedentes de hecho un párrafo si la opción es «Sí» y otro si la opción es «No».

A continuación, tenemos los hechos probados, la prueba practicada y las opciones más importantes, a saber, la «Antigüedad», el «Salario» y la «Relación de retrasos». Estas 3 entradas serán las que determinen el resto de la sentencia, tanto en su contenido como en su sentido estimatorio o desestimatorio. La IA estaría encargada de varias funciones en la presente aplicación. Las variables como nombre, número de procedimiento, hechos probados, etc. serían valores introducidos por el usuario. Las variables serían las siguientes:

Una vez que el usuario cumplimentara todos los datos, las tareas a realizar por la IA serían las siguientes:

- Calcular el promedio del retraso en el abono de los salarios con los datos introducidos en «Relación de los retrasos». Esta relación de los retrasos serían hechos probados.

Imagen 2. Imagen de los cuadros a rellenar por el usuario

Formulario de datos para el usuario:

- Fecha: 14/9/2023
- Número de juicio: []
- Nombre del demandante: []
- Nombre de la institución demandada: []
- ¿Comparecen ambas partes?: Sí No
- Hechos probados: []
- Pruebas:
 - Documental
 - Testifical
 - Interrogatorio dte
 - Pericial
 - Forense
 - Interrogatorio ddo
- Antigüedad: 14/9/2023
- Salario día: []
- Relación de retrasos: []
- Botones: Guardar, Resolver

Fuente: elaboración propia a partir de la aplicación.

- Calcular la mediana del retraso en el abono de los salarios.
- Redactar un razonamiento sobre por qué los retrasos son graves (para lo que se toma la media) y continuados (para lo que se toma la mediana).
- En caso de contemplar que los retrasos son graves y continuados, rellenará con todos los datos facilitados por el usuario el modelo de sentencia estimatoria, en caso contrario utilizará el modelo de sentencia desestimatoria.
- Calcular la indemnización por extinción, a la fecha de la sentencia, en caso de estimación de la demanda.

Como ya se ha señalado, esta primera aplicación parte de un modelo tipo AlphaGo, es decir, cuyo entrenamiento parte del conocimiento de millones de partidas previas. En el caso del AlphaJudge, lo que hacemos es alimentar al modelo del lenguaje, la IA, con jurisprudencia sobre el retraso en el abono del salario.

Para interactuar con GPT-4, utilizamos prompts², sin embargo, podemos modular la manera en la que GPT-4 nos contesta a través de instrucciones previas. Estas instrucciones previas son las que utilizaremos para alimentar a GPT-4 con la jurisprudencia sobre la extinción por retraso en el abono de salarios. En Python aparecería así:

```
response = openai.ChatCompletion.create(
    model="gpt-4",
    messages=[
        {"role": "system", "content": reglas},
        {"role": "user", "content": prompt},
    ]
)
texto = response['choices'][0]['message']['content']
```

Lo que se aprecia arriba es la diferenciación de roles. El *role system* es aquel en el que introducimos las reglas sobre cómo resolver los problemas que se planteen a GPT-4 y el

² Un «prompt» es una instrucción o estímulo que se proporciona a un modelo o sistema para obtener una respuesta o acción específica. En el contexto de los modelos de lenguaje, es el texto inicial que se introduce para que el modelo genere una continuación o respuesta basada en ese texto.



role user es aquel en el que se introduce el problema a resolver, a saber, la relación de retrasos. El cuadro de texto que observamos en la imagen 2, con el título «Relación de retrasos», es el prompt que se la dará a GPT-4 para que resuelva el problema conforme a las reglas dadas. Estas reglas se encuentran en un archivo de texto.

```
with open('resolutor.txt', 'r', encoding="utf8") as file:  
    reglas = file.read()
```

4.1.2. Instrucciones para la inteligencia artificial

Las reglas para resolver el pleito son la piedra fundamental en la configuración de la aplicación, dado que de la forma en la que se expliquen las reglas dependerá que la IA resuelva correctamente el pleito y nos ofrezca un razonamiento adecuado. La finalidad es que la IA haga una sentencia que bien pudiera haber hecho un juez humano, por lo que no basta con que resuelva estimar o desestimar, sino que debe motivar la razón por la cual considera que el retraso es grave, leve, continuado o esporádico. Las reglas contenidas en el archivo de texto son las siguientes:

Como juez especializado, se le solicita emitir un fallo basado en la gravedad de los retrasos salariales y su justificación para la extinción del contrato de trabajo. Utilice únicamente la siguiente jurisprudencia:

TS (22/12/2008): retrasos > 11.20 días en 336 días son graves.

TS (10/01/2023): retraso promedio de 10.5 días en un año es referencia.

TSJ Madrid (26/09/2019): retrasos breves/esporádicos no justifican terminación.

Instrucciones:

Calcule retraso promedio y mediano de todos los pagos. Los salarios suelen pagarse entre el 1 y el 5 del mes siguiente.

Presente: «Retraso medio: X días. Mediana: Y días».

Relacione promedio y mediana matemáticamente y su relevancia en este caso.

Con base en la jurisprudencia, determine si los retrasos son graves. Cite las sentencias.

Concluya con «condenar» si justifica extinción o «absolver» si no.

En estas reglas se hace uso de varias técnicas de *prompt engineering* (White *et al.*, 2023). Básicamente, la estructura a seguir es la siguiente:

- Sujeto a simular: «Como juez especializado».

El modelo se coloca en el papel de un juez especializado, lo que establece el tono y la perspectiva desde la cual debe abordar la tarea.

- Tarea: «se le solicita emitir un fallo basado en la gravedad de los retrasos salariales y su justificación para la extinción del contrato de trabajo».

Se define claramente lo que se espera que haga el modelo: emitir un fallo basándose en la gravedad de los retrasos salariales.

Pasos para completar la tarea:

- «Calcule retraso promedio y mediano de todos los pagos».
- «Relacione promedio y mediana matemáticamente y su relevancia en este caso».
- «Con base en la jurisprudencia, determine si los retrasos son graves. Cite las sentencias».
- «Concluya con "condenar" si justifica extinción o "absolver" si no».

Estos pasos guían al modelo a través de la tarea, asegurando que aborde todos los aspectos necesarios para llegar a una conclusión informada.

- Contexto/restricciones: «Utilice únicamente la siguiente jurisprudencia: [listado de sentencias]».

Se proporciona contexto sobre las sentencias judiciales relevantes y se establece una restricción para que el modelo solo utilice esa información específica al tomar una decisión.

- Objetivo: «Emitir un fallo bien fundamentado y respaldado sobre la gravedad de los retrasos».

El objetivo final es que el modelo proporcione un fallo basado en la información y las instrucciones proporcionadas.

- Formato de salida: «Presente: "Retraso medio: X días. Mediana: Y días"» y «Concluya con "condenar" o "absolver"».

Se especifica cómo debe presentarse la respuesta, lo que ayuda a obtener una salida clara y coherente.

Lo que tenemos aquí es la estructura básica del prompt, a saber: sujeto a simular, tarea, pasos para completar la tarea, contexto/restricciones, objetivo y formato de salida.

Con esta instrucción previa con la que se alimenta a la IA y su inclusión dentro de la aplicación, podemos ya comenzar a ofrecer relaciones de retrasos en el abono de los salarios para que comience a analizar los datos, sacar conclusiones y razonar en consecuencia.

Como elementos esenciales para clarificar el modelo, ha de explicarse que la expresión «Concluya con "condenar" o "absolver"» responde a que la aplicación previamente está configurada para que si la última frase de la respuesta de la IA contiene las palabras «condenar» o «absolver», elegirá el modelo de conclusión de sentencia adecuado, completando los nombres de las partes y, más importante, efectuando el cálculo de la indemnización por extinción (en la fecha de la sentencia) en caso de condena.

```
for frase in frases:
    if "CONDENAR" in frase:
        numero = 0
    elif "ABSOLVER" in frase:
        numero = 1
    else:
        respuesta_modificada += frase + "."
```

Otra cuestión necesaria que explicar es por qué se calcula la mediana además del promedio. La jurisprudencia tiene en cuenta siempre el promedio en los retrasos a la hora de resolver sobre la extinción: un retraso promedio de 11,20 días en 336 días es grave (Sentencia del Tribunal Supremo –STS– de 22 de diciembre de 2008, rec. 294/2008) y un retraso promedio de 10,50 días tomando como referencia 1 año también es grave (STS de 10 de enero de 2023, rec. 2166/2021). Sin embargo, el retraso ha de ser continuado y no esporádico, de ahí que se tome el cálculo de la mediana para apreciar si efectivamente los retrasos habidos responden o no a una tendencia³.

4.1.3. Pruebas y evaluación del programa

Expuesto todo lo que antecede, ya solo queda mostrar ejemplos de cómo, con una relación de retrasos, razona la IA.

³ La mediana puede ofrecer información sobre si los retrasos son esporádicos o continuados. Así, si la mediana es cercana a cero o muy baja, pero la media es alta, esto podría indicar que, aunque la mayoría de los meses no tienen retrasos significativos (o tienen retrasos cortos), hay algunos meses con retrasos extremadamente largos. Estos retrasos extremos serían esporádicos pero significativos. Si la mediana es alta y cercana a la media, esto sugiere que los retrasos son consistentes y continuados a lo largo del tiempo. No hay muchos valores extremos que distorsionen la media.

Como se puede observar en la imagen 3, una vez introducidos los datos, la vista previa del modelo se actualiza mostrando la argumentación en su FJ 4.º y el fallo, estimatorio con el cálculo de la indemnización.

Imagen 3. Aplicación AlphaJudge resolviendo un pleito

Rellenador de modelo

Fecha: 10/10/2023

Número de juicio: 123/2023

Nombre del demandante: D. Pedro Pérez

Nombre de la institución demandada: EMPRESA, SA

¿Comparecen ambas partes?
 Sí
 No

Hechos probados: El trabajador...

Pruebas:
 Documental Testifical
 Pericial Forense
 Interrogatorio dte. Interrogatorio ddo.

Antigüedad: 13/09/2018

Salario día: 12,34

Relación de retrasos:
-Mensualidad de marzo de 2021 se le abona el 14.04.2021.
-Mensualidad de abril de 2021 se le abona el 02.06.2021.
-Mensualidad de mayo de 2021 se le abona el ...

Resolver **Guardar**

En atención a los datos que obran en el relato fáctico, se computan los días de retrasados en los pagos sobre la fecha estipulada (es decir, el día 5 del mes siguiente). Dado que se mencionan 20 meses, es posible calcular estas cantidades.

El retraso medio en el abono de los salarios ha sido aproximadamente de 13 días al tomar en cuenta todos los meses mencionados. Así mismo, al arreglar los datos en orden creciente y hallar el valor central conseguimos la mediana, cuyo resultado sería de 12 días.

La relación matemática entre el promedio y la mediana nos muestra una visión clara de la distribución y tendencia central de los datos. Cuando ambos valores son similarmente altos, como en este caso, indica que la mayoría de los pagos se demoraron más de lo esperado y sin ser esporádicos, lo cual ofrece una base sólida para nuestra decisión.

Ciertamente, no todos los retrasos son motivo para solicitar la terminación del contrato. Según la sentencia del TSJ de Madrid del 26 de septiembre de 2019, tienen que ser retrasos que no sean breves ni esporádicos. No obstante, aquí los datos indican una demora continua y significativa, lejos de ser breve y esporádica.

Además, desde el punto de vista de la jurisprudencia del TS, los retrasos se consideran graves cuando superan un promedio de 11.20 días en un lapso de 336 días (sentencia de 22 de diciembre de 2008), y la sentencia del 10 de enero de 2023 mantiene un punto de referencia de un retraso de 10.5 días dentro de un año. En este caso, ambos umbrales se han sobrepasado claramente,

[...]

FALLO

Que DEBO ESTIMAR Y ESTIMO, la demanda interpuesta por D. Pedro Pérez contra EMPRESA, SA, y FOGASA y por ende debo declarar y declaro extinguida la relación laboral que le unía a , con efectos desde esta resolución judicial, condenando a la empresa a abonarle en concepto de indemnización la cantidad de 2.070,03 euros.

[...]

Fuente: elaboración propia a partir de la aplicación.

Para una mayor claridad vamos a plantear 3 supuestos y enlazar únicamente la argumentación arrojada por la IA.

Relación de retrasos – Grave y continuados

- Mensualidad de marzo de 2021 se le abona el 14 de abril 2021.
- Mensualidad de abril de 2021 se le abona el 2 de junio de 2021.
- Mensualidad de mayo de 2021 se le abona el 21 de junio de 2021.
- Mensualidad de junio de 2021 se le abona el 9 de julio de 2021.
- Mensualidad de julio de 2021 se le abona el 13 de agosto de 2021.
- Mensualidad de agosto de 2021 se le abona el 15 de septiembre de 2021.
- Mensualidad de septiembre de 2021 se le abona el 5 de octubre de 2021.
- Mensualidad de octubre de 2021 se le abona el 8 de noviembre de 2021.
- Mensualidad de noviembre de 2021 se le abona el 3 de diciembre de 2021.
- Mensualidad de diciembre de 2021 se le abona el 17 de enero de 2022.
- Mensualidad de enero de 2022 se le abona el 7 de febrero de 2022.
- Mensualidad de febrero de 2022 se le abona el 17 de marzo de 2022.
- Mensualidad de marzo de 2022 se le abona el 7 de abril de 2022.
- Mensualidad de abril de 2022 se le abona el 6 de mayo de 2022.
- Mensualidad de mayo de 2022 se le abona el 13 de junio de 2022.
- Mensualidad de junio de 2022 se le abona el 25 de julio de 2022.
- Mensualidad de julio de 2022 se le abona el 27 de agosto de 2022.
- Mensualidad de agosto de 2022 se le abona el 30 de septiembre de 2022.
- Mensualidad de septiembre de 2022 se le abona el 18 de noviembre de 2022.
- Mensualidad de octubre de 2022 se le abona el 17 de diciembre de 2022.

Argumentación de la IA

En atención a los datos que obran en el relato fáctico, los retratos en los salarios presentan cierta recurrencia temporal. Para establecer la gravedad de estos retrasos procedemos a calcular, en primer lugar, el retraso medio, que resulta al sumar todos los retrasos y dividirlos por el número total de pagos. Haciendo este cálculo, obtenemos que el retraso medio en el abono de los salarios ha sido de 12.15 días.

Asimismo, para obtener la mediana, ordenamos de menor a mayor el número de días de retraso y buscamos el valor situado en medio. Si el número de pagos fuera par, se tomaría la media de los 2 valores centrales. No obstante, en este caso, al contar con un número impar de pagos, el cálculo resulta en una mediana de 12 días.





El promedio y la mediana nos dan medidas de la tendencia central de los datos. Son conceptos estrechamente relacionados, pero mientras el promedio aritmético equilibra todos los datos, la mediana se enfoca en el dato del medio, siendo menos sensible a los valores extremos. Por lo cual, la mediana es especialmente útil para observar si los valores están dispersos o si existen irregularidades o picos extremos en los datos. En este caso, con un promedio y mediana muy cercana, se sugiere una consistencia y continuidad en los retrasos, más que esporadicidad o presencia de valores extremos.

Analizando el escenario con la jurisprudencia aplicable, vemos que el retraso medio y mediano superan el límite de los 11.20 días estipulado en la STS del 22 de diciembre de 2008, así como también rebasan el umbral de los 10.5 días contemplado por el TS el 10 de enero de 2023.

Por otro lado, debemos tener en cuenta la Sentencia del Tribunal Superior de Justicia (STSJ) de Madrid de 26 de septiembre de 2019, la cual hace énfasis en que no se puede solicitar la terminación debido a retrasos breves y/o esporádicos. En este caso, aunque los retrasos no son breves –dado que superan los días promedios estipulados por el TS– tampoco son esporádicos, sino que revelan cierta persistencia.

Por tanto, teniendo en cuenta la consistencia en los retrasos, su gravedad expresada tanto en promedio como en mediana y con base en la jurisprudencia aplicable, debemos concluir que estos retrasos son graves y justifican la acción de extinción del contrato de trabajo por voluntad de la persona trabajadora.

Relación de retrasos – Breves y continuados

- Mensualidad de marzo de 2021 se le abona el 6 de abril de 2021.
- Mensualidad de abril de 2021 se le abona el 7 de mayo de 2021.
- Mensualidad de mayo de 2021 se le abona el 6 de junio de 2021.
- Mensualidad de junio de 2021 se le abona el 7 de julio de 2021.
- Mensualidad de julio de 2021 se le abona el 6 de agosto de 2021.
- Mensualidad de agosto de 2021 se le abona el 7 de septiembre de 2021.
- Mensualidad de septiembre de 2021 se le abona el 6 de octubre de 2021.
- Mensualidad de octubre de 2021 se le abona el 7 de noviembre de 2021.
- Mensualidad de noviembre de 2021 se le abona el 6 de diciembre de 2021.
- Mensualidad de diciembre de 2021 se le abona el 7 de enero de 2022.
- Mensualidad de enero de 2022 se le abona el 6 de febrero de 2022.
- Mensualidad de febrero de 2022 se le abona el 7 de marzo de 2022.
- Mensualidad de marzo de 2022 se le abona el 6 de abril de 2022.
- Mensualidad de abril de 2022 se le abona el 7 de mayo de 2022.
- Mensualidad de mayo de 2022 se le abona el 6 de junio de 2022.
- Mensualidad de junio de 2022 se le abona el 7 de julio de 2022.
- Mensualidad de julio de 2022 se le abona el 6 de agosto de 2022.





- Mensualidad de agosto de 2022 se le abona el 7 de septiembre de 2022.
- Mensualidad de septiembre de 2022 se le abona el 6 de octubre de 2022.
- Mensualidad de octubre de 2022 se le abona el 7 de noviembre de 2022.

Argumentación de la IA

En atención a los datos que obran en el relato fáctico, el retraso medio en el abono de los salarios ha sido de 1.05 días (12.6 días anuales), asimismo podemos calcular la mediana, cuyo resultado sería de 1.05 días.

La media y la mediana son los indicadores de tendencia central que nos dan un resumen de los datos. La media es la suma de todos los valores dividida por el número de valores, mientras que la mediana es el valor que divide a los datos en 2 partes iguales, es decir, el 50 % de los datos son menores que la mediana y el otro 50 % son mayores. En este caso, tanto la media como la mediana son de 1.05 días, lo que indica que la distribución de los retrasos es uniforme y centrada en este valor.

Basándome en las SSTS de 22 de diciembre de 2008 y 10 de enero de 2023, se consideran «graves» aquellos retrasos que superan un promedio de 10.5-11.20 días al año. Tanto el promedio como la mediana de retrasos en el presente caso están por debajo de este umbral, lo que indica que los retrasos podrían no ser graves.

Además, según la STSJ de Madrid del 26 de septiembre de 2019, los retrasos breves y/o esporádicos no podrían justificar una acción de terminación. En este caso, tanto la media como la mediana indican que los retrasos han sido constantes y no esporádicos, aunque breves (1.05 días).

En conclusión, aunque los retrasos han sido constantes, la duración de estos no alcanza la gravedad requerida por las sentencias del TS para justificar la extinción del contrato de trabajo por voluntad del trabajador.

Relación de retrasos – Esporádicos

- Mensualidad de marzo de 2021 se le abona el 3 de abril de 2021.
- Mensualidad de abril de 2021 se le abona el 2 de mayo de 2021.
- Mensualidad de mayo de 2021 se le abona el 10 de junio de 2021.
- Mensualidad de junio de 2021 se le abona el 4 de julio de 2021.
- Mensualidad de julio de 2021 se le abona el 5 de agosto de 2021.
- Mensualidad de agosto de 2021 se le abona el 9 de septiembre de 2021.
- Mensualidad de septiembre de 2021 se le abona el 3 de octubre de 2021.
- Mensualidad de octubre de 2021 se le abona el 4 de noviembre de 2021.
- Mensualidad de noviembre de 2021 se le abona el 8 de diciembre de 2021.
- Mensualidad de diciembre de 2021 se le abona el 4 de enero de 2022.
- Mensualidad de enero de 2022 se le abona el 5 de febrero de 2022.





- Mensualidad de febrero de 2022 se le abona el 3 de marzo de 2022.
- Mensualidad de marzo de 2022 se le abona el 10 de abril de 2022.
- Mensualidad de abril de 2022 se le abona el 3 de mayo de 2022.
- Mensualidad de mayo de 2022 se le abona el 4 de junio de 2022.
- Mensualidad de junio de 2022 se le abona el 5 de julio de 2022.
- Mensualidad de julio de 2022 se le abona el 9 de agosto de 2022.
- Mensualidad de agosto de 2022 se le abona el 3 de septiembre de 2022.
- Mensualidad de septiembre de 2022 se le abona el 4 de octubre de 2022.
- Mensualidad de octubre de 2022 se le abona el 5 de noviembre de 2022.

Argumentación de la IA

En atención a los datos que obran en el relato fáctico, el retraso medio en el abono de los salarios ha sido de 7.15 días, calculado sobre todos los pagos facilitados. Además, el retraso mediano es de 5 días.

La relación matemática entre el promedio y la mediana es esencial para comprender la distribución y tendencia central de los datos de retraso en los pagos salariales. El promedio es la suma de todos los retrasos dividida por el número total de salarios. Por otro lado, la mediana es el valor central en un conjunto de datos, lo que significa que el 50 % de los valores están por debajo de este punto y el 50 % están por encima. Al comparar el retraso promedio con el retraso mediano, podemos determinar si los retrasos son principalmente esporádicos o continuos.

Según la jurisprudencia aplicable al presente caso, los retrasos en los pagos salariales deben considerarse graves para justificar la terminación del contrato de trabajo por voluntad de la persona trabajadora. El TS, en su Sentencia de 22 de diciembre de 2008, establece que los retrasos que superen un promedio de 11.20 días en un lapso de 336 días se consideran «graves». Además, la Sentencia de 10 de enero de 2023 establece que un retraso promedio de 10.5 días dentro de 1 año sirve como punto de referencia.

Considerando estos criterios jurisprudenciales, los retrasos en los pagos salariales presentados en el caso no se consideran graves. Los retrasos promedio y mediano de 6.4 días y 4 días respectivamente se encuentran por debajo del límite establecido por el TS. Además, el retraso mediano de 4 días sugiere que los retrasos son más esporádicos que continuos. Además, al analizar la frecuencia, se evidencia que los retrasos son esporádicos, ya que solo en 4 ocasiones de los 20 pagos analizados se ha producido un retraso efectivo.

En vista de lo anterior y conforme a la jurisprudencia aplicable, los retrasos en los pagos salariales no pueden considerarse graves y continuados, sino más bien esporádicos. Por lo tanto, se concluye que la acción de extinción del contrato de trabajo por voluntad de la persona trabajadora no está justificada.

En atención a los datos ofrecidos al modelo, las respuestas, si bien podrían adoptar formas diversas en atención al estilo jurídico y de redacción de cada uno, sin embargo, ofrecen un razonamiento adecuado y una relación clara con la jurisprudencia facilitada.

Lo expuesto no es más que un ejemplo simple, la ventana de contexto del *role system* es limitada, sin embargo, mediante *fine-tuning*⁴ podrían crearse agentes de IA especializados en los distintos tipos de procedimiento (despidos, reclamación de cantidad, Seguridad Social, etc.). Bastaría con preentrenar estos modelos con jurisprudencia abundante y adecuada para poder resolver pleitos conforme a la doctrina actual e imperante (Peng *et al.*, 2023).

El modelo expuesto solo bebe de 3 sentencias, una de las cuales ni siquiera sería jurisprudencia *ex artículo 1.6 del Código Civil (la STSJ)* y, sin embargo, con 3 únicos datos resumidos y sin argumentación jurídica, la aplicación es capaz de detectar los elementos necesarios, hacer los cálculos, resolver en consecuencia y motivar su razonamiento.

4.2. AlphaZero Judge

Para este segundo enfoque del concepto de juez robot hemos tomado un caso concreto de resolución judicial, a saber, el de la impugnación de una resolución del Instituto Nacional de la Seguridad Social que revisa la situación de incapacidad del beneficiario y rebaja la calificación por mejoría. El artículo 200.2 de la Ley general de la Seguridad Social dispone que toda resolución, inicial o de revisión, por la que se reconozca el derecho a las prestaciones de incapacidad permanente (IP), en cualquiera de sus grados, hará constar necesariamente el plazo a partir del cual se podrá instar la revisión por agravación o mejoría del estado incapacitante profesional. La STS de 26 de marzo de 1987 señala que la revisión del grado de invalidez permanente por agravación o mejoría de la trabajadora presupone siempre una confrontación entre dos situaciones de hecho: la que determinó la primitiva declaración de invalidez y la existente cuando se lleva a efecto la revisión. Igualmente, la STS de 22 de diciembre de 2009 (rec. 2066/09) señala que la «mejoría» que justifica la revisión exige conceptualmente no solo comparar dos situaciones patológicas (la que determinó la declaración de IP y la existente cuando se lleva cabo la revisión) y llegar a la conclusión de que ha variado el cuadro de dolencias, sino –sobre todo– que esta variación tiene trascendencia cualitativa en orden a la capacidad de trabajo del declarado en IP, en tanto que alcance a justificar la modificación del grado reconocido, de forma tal que si las secuelas permanecen sustancialmente idénticas no hay cauce legal para modificar la calificación en su día efectuada.

En suma, este tipo de procedimientos se rige por una regla sencilla, a saber, para revisar la IP debido a agravación o mejoría, se debe comparar la situación inicial de invalidez

⁴ El *fine-tuning* en grandes modelos de lenguaje se refiere al proceso de adaptar un modelo previamente entrenado en un conjunto de datos grande y general (como todo el texto de internet) a un conjunto de datos más pequeño y específico. Este proceso permite que el modelo se especialice en tareas particulares o se ajuste a instrucciones específicas.

con la situación actual. Si las condiciones no han cambiado significativamente, no se puede modificar la calificación original.

Como ya se indicó, el modelo AlphaZero fue instruido únicamente con las reglas del Go, el ajedrez y el shogi, perfeccionando su entrenamiento jugando contra sí mismo, pero lo decisivo es el modelo de entrenamiento, es decir, no jugadas humanas previas, sino simplemente las reglas con un aprendizaje por refuerzo contra sí mismo.

En el presente caso, únicamente se indicará a la IA la regla antedicha y, con la misma, deberá resolver si las situaciones médicas previas y actuales suponen una mejora o agravamiento.

GPT-4, al igual que otros grandes modelos de lenguaje (LLM), ha demostrado capacidades notables en la comprensión y generación de lenguaje natural en diversas materias, entre ellos, la medicina. La medicina es el punto esencial a la hora de resolver el tipo de procedimiento sobre el que va a versar la aplicación. Así, GPT-4 ha sido evaluado para el United States Medical Licensing Examination (USMLE), un programa de examen en 3 etapas utilizado para evaluar la capacidad de un médico para aplicar conocimientos, conceptos y principios, demostrando sus habilidades fundamentales y con su superación otorgar licencias en los Estados Unidos. Además, se evaluó su rendimiento en el conjunto de datos MultiMedQA⁵. Los resultados mostraron que GPT-4 superaba la puntuación de aprobado en el USMLE y superaba a modelos anteriores y a modelos específicamente ajustados en conocimiento médico. Se investigó la influencia de las preguntas de prueba que contienen texto e imágenes en el rendimiento del modelo y se estudió la calibración de probabilidad, que es de vital importancia en aplicaciones de alto riesgo como la medicina. Además, se exploró la capacidad de GPT-4 para explicar el razonamiento médico, personalizar explicaciones a los estudiantes y elaborar interactivamente nuevos escenarios contrafactuales en torno a un caso médico (Nori *et al.*, 2023).

En definitiva, estamos ante una regla sencilla cuya aplicación precisa o bien de peritos, o bien de un conocimiento extenso de materias médicas. GPT-4 se presenta en este escenario como una opción competente, dado sus conocimientos médicos y su capacidad para efectuar razonamientos.

AlphaZero, con su capacidad para aprender sin datos previos, representa un enfoque más adaptativo y autónomo. En el contexto judicial, esto podría traducirse en un juez robot que aprende y se adapta con cada caso que evalúa, ajustando y refinando su comprensión de la ley a medida que se enfrenta a nuevos escenarios y desafíos (Branting, 2017).

⁵ Un modelo del lenguaje (LLM) para fines médicos que combina HealthSearchQA y 6 conjuntos de datos actuales de respuesta a preguntas abiertas publicada el 26 de diciembre de 2022 por Google Research y Deepmind.

4.2.1. Aplicación

La aplicación creada parte de un modelo de sentencia, debiendo rellenar los valores no repetitivos, a saber, número de procedimiento, nombres, hechos probados, etc. Siendo la interfaz gráfica la siguiente:

Imagen 4. Imagen global de la aplicación AlphaZero Judge

The screenshot displays the 'Rellenador de modelo' (Model Filler) application window. The interface is divided into two main sections: a form for inputting case details on the left and a text area for the generated sentence on the right.

Form Section (Left):

- Fecha:** A date picker set to 16/9/2023.
- Número de juicio:** An empty text input field.
- Nombre del demandante:** An empty text input field.
- Nombre de la institución demandada:** An empty text input field.
- Options:** Three checkboxes for 'INSS', 'TGSS', and 'Mutua', all currently unchecked.
- Hechos probados:** An empty text input field with '+' and '-' buttons below it.
- Incapacidad permanente revisada:** An empty text input field.
- Pruebas:** A grid of checkboxes for 'Documental', 'Testifical', 'Perito actor', 'Perito ddo.', 'Detect. ddo.', and 'Fortense', all unchecked.
- Patologías y limitaciones anteriores:** A dropdown menu set to '2023' and an empty text input field.
- Patologías y limitaciones actuales:** A dropdown menu set to '2023' and an empty text input field.
- Buttons:** 'Guardar' (Save) and 'Resolver' (Solve) buttons at the bottom.

Sentence Text Area (Right):

SENTENCIA

En Las Palmas de Gran Canaria, a [Fecha].

Visto por mí, D. Javier Ercilla García, Magistrado del Juzgado de lo Social n.º 10 de Las Palmas de Gran Canaria, en audiencia pública, el juicio sobre Resolución del contrato de trabajo, seguido ante este juzgado bajo n.º [Número de juicio], promovido a instancia de [Nombre del demandante], contra [Nombre de la institución demandada] y FOGASA, atendiendo a los siguientes:

ANTECEDENTES DE HECHO

PRIMERO. La parte actora arriba indicada presentó en el decanato una demanda que fue repartida a este juzgado y en la que, previa alegación de los hechos y fundamentos de derecho que estimó procedentes, terminó suplicando que se dictara una sentencia de conformidad con sus pretensiones.

SEGUNDO. Admitida a trámite la demanda, las partes fueron citadas al acto de juicio. Comparecidas las partes, asistidas en la forma que consta en el acta, se pasó al acto de juicio. En él, y una vez que se hubo efectuado la dación de cuenta de los antecedentes, la parte actora se ratificó en su demanda. La parte demandada se opuso a la demanda y la contestó formulando las alegaciones que constan en acta. [...].

TERCERO. En la sustanciación de estos autos se han observado todas las prescripciones legales aplicables salvo los plazos para dictar sentencia.

HECHOS PROBADOS

XXXX. Se agotó la vía previa.

FUNDAMENTOS DE DERECHO

PRIMERO. A los efectos del art. 97.2 Ley Reguladora de la Jurisdicción Social (LRJS), conviene resaltar que los hechos declarados se ha establecido en base a la valoración conjunta de la actividad probatoria desarrollada, en concreto de [prueb

Fuente: elaboración propia a partir de la aplicación.

En la imagen se puede observar la misma estructura que en la anterior, a saber, una vista previa de la sentencia que se está dictando a la derecha y a la izquierda las distintas entradas, de manera que, conforme se escribe en los recuadros de la izquierda, la sentencia se va actualizando con todos los detalles.

El elemento diverso en la aplicación se refiere a las patologías y limitaciones a comparar en el tiempo:

Imagen 5. Cuadros de comparación de patologías y limitaciones

Incapacidad permanente revisada

Pruebas

Documental Testifical

Perito actor Perito ddo.

Detect. ddo. Fortense

Patologías y limitaciones anteriores

Año 2023

Patologías y limitaciones actuales

Año 2023

Relación de retrasos

Resolver

Fuente: elaboración propia a partir de la aplicación.

La IA, en este caso, tendría encargadas las siguientes tareas:

- Analizar las patologías y limitaciones existentes al tiempo de la calificación de IP original y compararlas con las patologías y limitaciones existentes al tiempo.
- Redactar un razonamiento sobre si ha existido una mejoría en la situación patológica del beneficiario o si ha existido un empeoramiento o mantenimiento de la situación.
- En caso de contemplar que no ha habido mejoría, rellenará con todos los datos facilitados por el usuario el modelo de sentencia estimatoria, en caso contrario utilizará el modelo de sentencia desestimatoria.

4.2.2. Instrucciones para la inteligencia artificial

En este caso, la regla para resolver el pleito es muy sencilla, a saber, analizar ambas situaciones conforme a lo ingresado por el usuario (año y patologías) y determinar si ha habido o no mejoría. Las reglas contenidas en el archivo de texto son las siguientes:

Como juez especializado, se le pide evaluar la revisión de una IP basada en la evolución de las patologías del demandante. Debe comparar la situación patológica inicial con la actual y determinar si ha habido mejoría, estabilidad o empeoramiento. Siga estas instrucciones:

- Compare la situación patológica inicial del trabajador cuando fue declarado incapacitado con la situación actual.
- Determine si la mejoría es lo suficientemente relevante como para cambiar el grado de incapacidad.
- Concluya con «condenar» si no hay mejoría o «absolver» si hay mejoría significativa.

Como ya se expuso antes, la estructura para seguir ha sido la siguiente:

- Sujeto a simular: «Como juez especializado».
El modelo adopta el papel de un juez especializado, estableciendo la perspectiva y el tono desde el cual debe abordar la tarea.
- Tarea: «se le pide evaluar la revisión de una incapacidad permanente basada en la evolución de las patologías del demandante».
Se define claramente lo que se espera que haga el modelo: evaluar la revisión de una incapacidad basándose en la evolución de las patologías.
- Pasos para completar la tarea:
 - «Compare la situación patológica inicial del trabajador cuando fue declarado incapacitado con la situación actual».
 - «Determine si la mejoría es lo suficientemente relevante como para cambiar el grado de incapacidad».
 - «Concluya con "condenar" si no hay mejoría o "absolver" si hay mejoría significativa».

Estos pasos guían al modelo a través de la tarea, asegurando que aborde todos los aspectos necesarios para llegar a una conclusión informada.

- Contexto/restricciones: no se especifican restricciones adicionales en el prompt, pero se entiende que el modelo debe basar su evaluación en la evolución de las patologías del demandante.
- Objetivo: «evaluar la revisión de una incapacidad permanente basada en la evolución de las patologías del demandante y determinar si ha habido mejoría, estabilidad o empeoramiento».

El objetivo final es que el modelo proporcione una evaluación basada en la información y las instrucciones proporcionadas.

- Formato de salida: «Concluya con "condenar" si no hay mejoría o "absolver" si hay mejoría significativa».

Se especifica cómo debe presentarse la respuesta, lo que ayuda a obtener una salida clara y coherente.

4.2.3. Pruebas y evaluación del programa

Expuesto todo lo que antecede, ya solo queda mostrar ejemplos de cómo, con una relación de patologías previas y actuales, razona la IA.

Como se puede observar en la imagen 6, una vez introducidos los datos, la vista previa del modelo se actualiza mostrando la argumentación en su FJ 5.º y el fallo estimatorio con el reconocimiento de la IP total.

Imagen 6. Aplicación AlphaZero Judge resolviendo un pleito

The screenshot shows a web application window titled "Rellenador de modelo". On the left, there is a form with the following fields and options:

- Fecha: 18/10/2023
- Número de juicio: 123/2023
- Nombre del demandante: D. Pedro Pérez
- Nombre de la institución demandada: INSS, TGSS, Mutua
- Hechos probados: El actor... (with + and - buttons)
- Incapacidad permanente revisada: Incapacidad Permanente Total
- Pruebas: Documental, Testifical, Perito actor, Perito ddo., Detect. ddo., Fortense

On the right, a preview window shows the generated text:

Así las cosas es necesario destacar que la revisión de la invalidez exige necesariamente de un juicio comparativo entre la situación patológica del trabajador al tiempo de ser declarado como tal y la que se produce en el momento posterior, en concreto aquel en el que se pretende llevar a cabo la antes de dicha revisión, juicio comparativo a raíz del cual deben extraerse como presupuestos conjuntos y necesarios para que la misma pueda prosperar coma cuando lo sea por mejoría, por una parte la efectiva y real mejoría en el estado de salud del inválido punto así lo reconoce el Tribunal Supremo, por ejemplo, en Sentencia de 15 de enero y 26 de marzo de 1987, al señalar que «la revisión del grado de invalidez permanente por agravación o mejoría del trabajador, presupone siempre una confrontación entre 2 situaciones De hecho, la que determina la primitiva declaración de invalidez y la existente cuando se lleva a efecto la revisión, sin que la mera circunstancia de que concorra alguna de aquellas causas, determinan por sí solo la modificación del grado de incapacidad si la naturaleza de las dolencias tiene idéntica repercusión en la capacidad laboral del trabajador» punto el segundo lugar de su presupuesto necesario que la mejoría sea especial y significativamente relevante, hasta el punto de permitir la subsunción dos luego estado del supuesto derecho definidor del nuevo grado incapacitado de que se le reconozca.

◀

Patologías y limitaciones anteriores

Año

Cervicalgia en paciente con antecedentes de hernia discal C6-C7 intervenida (2-2008 discectomía C6-C7+Fusión). Espondiloartrosis lumbar

Patologías y limitaciones actuales

Año

Disectomía C6-C7 intervenida en 2018. Espondiloartrosis lumbar, proceso de raquis cervical con movilización espontánea y balance articular activo limitado en últimos

[...] Vistos los preceptos legales citados y demás de general observancia y por la autoridad que me confiere el art. 117 de la Constitución Española y 1 de la Ley Orgánica del Poder Judicial,

FALLO

Que DEBO ESTIMAR y ESTIMO la demanda interpuesta por D. Pedro Pérez frente INSS y TGSS, reconociéndose al actor el grado de Incapacidad Permanente Total, derivada de contingencia común, condenándose el Ente demandado a abonar al demandante de pensión, practicándose en ejecución de sentencia las compensaciones u opciones que procedan en atención a las prestaciones incompatibles que perciba.

Notifíquese la presente Resolución a las partes del legal forma, haciéndose saber al tiempo que contra la misma cabe recurso de Suplicación, para ante la Sala de lo Social del Tribunal Superior de Justicia de Canarias.

Fuente: elaboración propia a partir de la aplicación.

Para una mayor claridad, vamos a plantear dos supuestos y enlazar únicamente la argumentación arrojada por la IA.

Relación de patologías 2015-2018: Empeoramiento

Las patologías de la parte demandante en 2015 eran:

«Cervicalgia en paciente con antecedentes de hernia discal C4-C5 intervenida (10-2015 discectomía C4-C5 + Fusión). Espondiloartrosis torácica».

Las patologías de la parte demandante en 2023 eran:

«Cervicalgia crónica con radiculopatía secundaria a hernia discal recurrente en C4-C5. Espondiloartrosis torácica avanzada con inicio de compromiso neurológico. Estenosis del canal lumbar secundaria a espondiloartrosis. Limitación funcional para actividades que requieran esfuerzo físico, levantamiento de peso y movimientos repetitivos del tronco».

Argumentación de la IA

El presente caso requiere realizar un juicio comparativo entre la situación patológica del trabajador en el momento de ser declarado como incapacitado permanente y la situación patológica actual.





En 2015, la parte demandante presentaba una cervicalgia en un paciente con antecedentes de hernia discal C4-C5 intervenida y espondiloartrosis torácica. Por su parte, en 2023, la situación patológica del actor incluye una cervicalgia crónica con radiculopatía secundaria a hernia discal recurrente en C4-C5, espondiloartrosis torácica avanzada con inicio de compromiso neurológico y estenosis del canal lumbar secundaria a espondiloartrosis. Además, se indica que existe una limitación funcional para actividades que requieran esfuerzo físico, levantamiento de peso y movimientos repetitivos del tronco.

Analizando esta información, se observa una clara evolución de las patologías y limitaciones del trabajador. En primer lugar, se observa una recurrencia de la hernia discal en C4-C5, lo que sugiere que la intervención quirúrgica realizada en 2015 no fue completamente exitosa en términos de eliminar completamente el problema. Además, se añade una espondiloartrosis torácica avanzada con compromiso neurológico.

Estas nuevas patologías, sumadas a la estenosis del canal lumbar, indican un empeoramiento significativo en la situación patológica del trabajador. Asimismo, la limitación funcional para actividades que requieran esfuerzo físico, levantamiento de peso y movimientos repetitivos del tronco confirma la disminución de su capacidad laboral.

En consecuencia, se puede concluir de manera clara y fundamentada que la situación patológica global del actor ha empeorado desde el momento en que fue declarado como incapacitado permanente hasta el presente. La evolución de sus patologías, junto con la aparición de nuevas limitaciones, demuestra que su estado de salud ha sufrido un deterioro significativo.

Relación de patologías 2015-2018: Mejoramiento

Las patologías de la parte demandante en 2015 eran:

«Cervicalgia en paciente con antecedentes de hernia discal C4-C5 intervenida (10-2015 discectomía C4-C5 + Fusión). Espondiloartrosis torácica».

Las patologías de la parte demandante en 2023 eran:

«Antecedentes de cervicalgia post-intervención en C4-C5, actualmente asintomático. Espondiloartrosis torácica estable sin signos de progresión. Recuperación funcional completa con capacidad para realizar actividades laborales de moderada intensidad, incluyendo levantamiento ocasional de peso y tareas que requieran movilidad del tronco. Se recomienda evaluación periódica y mantenimiento de ejercicios de fortalecimiento».

Argumentación de la IA

En atención a las patologías acreditadas, podemos realizar un juicio comparativo entre la situación patológica del trabajador en el momento de ser declarado como incapacitado permanente en 2015 y la situación actual en 2023.

En 2015, el trabajador presentaba cervicalgia y antecedentes de hernia discal C4-C5 intervenida, además de espondiloartrosis torácica. Estas patologías, sin duda, limitaban la capacidad laboral del trabajador y fueron razón para su declaración de IP.





Sin embargo, en la situación actual en 2023 se observa una mejoría significativa en la condición del trabajador. Se indica que la cervicalgia post-intervención en C4-C5 se encuentra asintomática, lo que implica una mejoría evidente en comparación con la situación anterior. Además, se menciona que la espondiloartrosis torácica se encuentra estable y sin signos de progresión.

Es importante destacar que se menciona una recuperación funcional completa, lo que implica que el trabajador tiene la capacidad para realizar actividades laborales de moderada intensidad, incluyendo levantamiento ocasional de peso y tareas que requieran movilidad del tronco. Además, se recomienda una evaluación periódica y el mantenimiento de ejercicios de fortalecimiento.

Con base en esta información, podemos concluir que ha habido una mejoría significativa en la condición patológica del demandante desde la fecha en que se le declaró como incapacitado permanente en 2015. La cervicalgia post-intervención se ha vuelto asintomática y la espondiloartrosis torácica se encuentra estable. Además, se destaca que el demandante ha logrado una recuperación funcional completa con capacidad para realizar actividades laborales de moderada intensidad.

En consecuencia, concluyo que la situación patológica del demandante ha mejorado y que la revisión de la IP está justificada en este caso.

En atención a los datos ofrecidos al modelo, las respuestas, si bien podrían adoptar formas diversas en atención al estilo jurídico y de redacción de cada uno, sin embargo, ofrecen un razonamiento adecuado y una relación clara con la jurisprudencia facilitada.

5. Estructura y anatomía del programa

Imagen 7. Estructura del juez robot

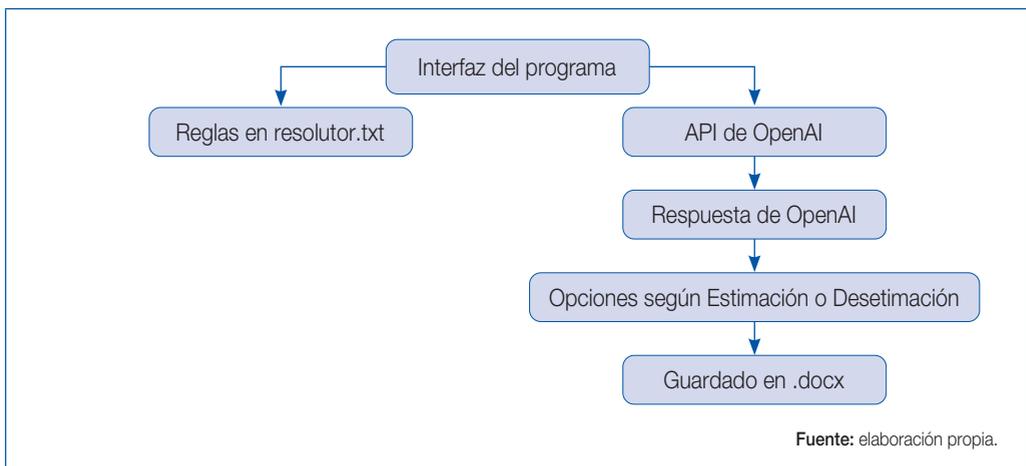


Imagen 8. Comunicación de los datos a GPT-4 y respuesta

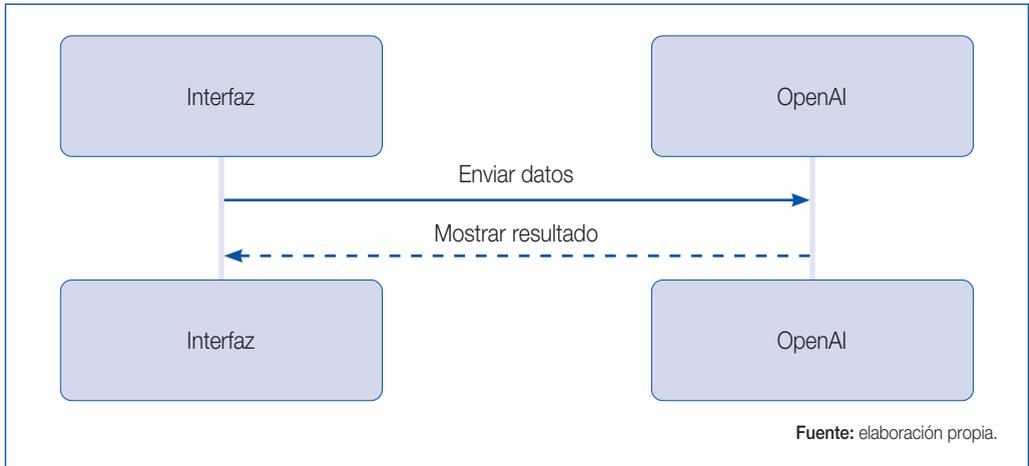


Imagen 9. Dinámica básica de la aplicación



6. Desafíos éticos y técnicos

La propuesta de implementar jueces robot en el sistema judicial, apoyados por modelos avanzados de IA, plantea una serie de desafíos éticos y técnicos que deben ser cuidadosamente considerados. Estos desafíos no solo se relacionan con la funcionalidad y precisión de tales sistemas, sino también con las implicaciones más amplias para la justicia, la equidad y la sociedad en general.

6.1. Desafíos éticos

En la consideración de la implementación de sistemas de IA en el ámbito judicial surgen diversas problemáticas. Al considerar el modelo AlphaJudge, se presupone que la IA adjudicaría litigios basándose en la jurisprudencia previamente establecida por órganos superiores, en particular, el TS. No obstante, esta premisa conlleva a varias interrogantes. Una de las principales es la viabilidad de que estos sistemas de IA suplanten completamente a los jueces humanos, dado que existirán casos sin jurisprudencia previa del TS. Estos se identifican como asuntos de interés casacional, es decir, aquellos que aborden puntos y cuestiones con jurisprudencia contradictoria o que apliquen normas recientes sin una doctrina jurisprudencial previa del TS. Adicionalmente, hay litigios que no admiten recurso más allá de la primera instancia y carecen de jurisprudencia, lo que, en teoría, los haría idóneos para ser resueltos por un sistema de IA. Sin embargo, la ausencia de jurisprudencia los excluye.

Un sistema AlphaJudge podría resultar en una homogeneización de las decisiones judiciales, pero también una cristalización del derecho. Si se anticipa el criterio de resolución, el papel del abogado se vería mermado, limitando la introducción de interpretaciones innovadoras o contrarias a la jurisprudencia existente. Es improbable que un AlphaJudge emita una sentencia contraria a la doctrina del TS, ya que su programación se basa en dicha doctrina. En lugar de aplicar el derecho de manera dinámica, el sistema simplemente verificaría la coincidencia de hechos con casos previamente resueltos.

Transparencia y explicabilidad: una preocupación ética primordial en la implementación de sistemas de IA es la claridad en el proceso decisorio. Los algoritmos de IA, particularmente aquellos fundamentados en redes neuronales profundas, frecuentemente funcionan como entidades opacas (Oswald *et al.*, 2018), lo que complica la comprensión de su metodología decisoria. En el ámbito judicial, donde las resoluciones pueden influir de manera determinante en el destino de los individuos, es imperativo que se pueda elucidar y fundamentar dichas decisiones.

Sesgo y discriminación: los algoritmos de IA se nutren de datos para su entrenamiento. Si estos datos reflejan prejuicios o discriminaciones arraigadas en la sociedad, emerge un riesgo tangible de perpetuar o incluso intensificar dichos sesgos (Chouldechova y Roth, 2018). A modo ilustrativo, si un sistema judicial automatizado se capacita con decisiones históricas que manifiestan prácticas discriminatorias, podría perpetuar resoluciones sesgadas en el futuro.

Autonomía y responsabilidad: la cesión de responsabilidades judiciales a un sistema automatizado suscita interrogantes en torno a la autonomía y la responsabilidad. Ante una resolución errónea o injusta por parte de un sistema judicial automatizado, surge la cuestión de determinar la responsabilidad (Rahwan *et al.*, 2019). ¿Recae en el equipo desarrollador del algoritmo, en los jueces que lo adoptaron o en el propio sistema? (Mittelstadt *et al.*, 2016).

Por otro lado, el modelo AlphaJudge Zero plantea un enfoque distinto, donde el sistema de IA aprendería y aplicaría las normas de manera autónoma. Este modelo invita a reflexionar sobre las implicaciones y el tipo de sistema judicial que la sociedad desea establecer.

6.2. Desafíos técnicos

Complejidad de la normativa jurídica: la legislación, por su naturaleza, es intrínsecamente compleja y susceptible a ambigüedades. A diferencia de un sistema con reglas determinadas y fijas, la interpretación y aplicación de la normativa jurídica demanda una comprensión detallada del contexto y las circunstancias particulares de cada situación. La creación de un algoritmo capaz de maniobrar a través de esta complejidad representa un reto técnico de magnitud (Ashley, 2017).

Adaptabilidad: las normas jurídicas y sus interpretaciones están en constante evolución. Es imperativo que un juez automatizado pueda adaptarse a las modificaciones legislativas y a las recientes interpretaciones judiciales. Esto implica un grado de flexibilidad y aprendizaje constante que trasciende los modelos convencionales de IA (Branting, 2017).

Seguridad y privacidad: los jueces automatizados tendrían acceso a un volumen considerable de información personal y confidencial. Es crucial garantizar la seguridad de estos datos y proteger la privacidad de las partes involucradas (Zarsky, 2013).

Interpretación y aplicación de la normativa: derivado de lo anteriormente expuesto, se ha deducido que un juez automatizado, basado exclusivamente en el conocimiento literal de las normativas (denominado AlphaJudge Zero), podría presentar limitaciones al abordar litigios que requieran una interpretación profunda de la norma. Esta conclusión se fundamenta en que ningún sistema jurídico es completamente cerrado y esta naturaleza abierta proviene de la complejidad inherente al comportamiento humano y de los teoremas propuestos por Gödel. La escuela de la Exégesis, que promovía una interpretación literal de las normas, fue superada por la realidad, que evidenció la necesidad de interpretar las normas para abordar adecuadamente los desafíos cotidianos.

Por ende, considerando que las normas operan como silogismos con premisas y conclusiones, es posible distinguir entre la premisa mayor, que podría ser «Todos los ladrones deben ser castigados», y la premisa menor, «Todos los ladrones». Al enfocarse en la premisa mayor, surge la necesidad de examinar cómo un juez automatizado podría interpretar las normas. Es inviable considerar un juez automatizado que actúe meramente como un transmisor de la ley. Por lo tanto, se plantea cómo este podría y si es factible que lo haga, llevar a cabo la labor intrínsecamente humana de interpretar una norma.

El TS, en sus sentencias, ha proporcionado directrices sobre la interpretación de las normas. Por ejemplo, en la Sentencia de 9 de junio de 2006 (rec. 3703/1999) se establece que la

interpretación de la ley implica la comprensión de su sentido y alcance. En otra Sentencia de 14 de febrero de 2008 (rec. 5490/2000), el TS destaca la importancia del criterio sistemático en la interpretación. Adicionalmente, en la Sentencia de 30 de junio de 2009 (rec. 304/2005), el TS resalta la relevancia de la realidad social como elemento interpretativo.

Estos criterios interpretativos deben ser herramientas esenciales para un juez automatizado. Sin embargo, la interpretación basada en la realidad social del momento en que se aplica la norma puede presentar desafíos significativos. Para abordar adecuadamente esta interpretación, es esencial que el juez automatizado esté equipado con una Teoría de la Mente avanzada, permitiéndole representar y comprender la realidad social. Manuel María Zorrilla Ruiz, expresidente del TSJ del País Vasco, ofreció una definición precisa de lo que se entiende por realidad social. Esta interpretación de la realidad social debe ser objetiva y distante para evitar la «ideologización judicial». Sin embargo, surge un desafío adicional: determinar quién proporcionará los datos necesarios para esta interpretación y cómo estos datos pueden influir en las decisiones judiciales.

En resumen, la principal distinción entre un juez humano y un juez automatizado radica en el determinismo inherente al último, en contraposición a la libertad de voluntad del primero. Es esencial considerar cuántos jueces automatizados individuales y distintos se crearían y qué criterios se establecerían para diferenciarlos. De lo contrario, se corre el riesgo de estandarizar las resoluciones judiciales.

7. Conclusiones

La idea de integrar la IA en el sistema judicial, específicamente en la figura de un juez robot, ha sido objeto de un intenso escrutinio y debate. A lo largo del presente ensayo, hemos explorado la evolución de modelos de IA como AlphaGo y AlphaZero, el concepto y potencial de los jueces robot, la aplicación práctica de ambos modelos y los desafíos éticos y técnicos asociados con su implementación. A medida que concluimos, es esencial resumir y reflexionar sobre los puntos clave y las implicaciones futuras de esta intersección entre tecnología y justicia.

Así pues, fuera de la espectacularidad que podría resultar de dos aplicaciones como las anteriores, lo cierto es que la IA únicamente ha llevado a cabo una labor lógica de analogía. Es el usuario, juez humano, quien ha introducido los «hechos probados» y con base en ellos se ha resuelto. La resolución ha respondido, ora a una regla jurisprudencial, ora a una regla legal, pero en todo caso la IA se ha limitado a efectuar un silogismo: «si es A, aplico la regla B y obtengo la solución C». Nos encontramos ante un robot juez determinista, cuya actuación dependerá de los «hechos probados» que se le introduzcan, por lo que la labor humana, del juez humano, sigue siendo imprescindible, dado que la valoración de la prueba y la consideración de qué es la realidad judicial, sigue siendo una tarea humana. De tal manera que las ambigüedades, contradicciones y umbrales de duda que una prueba pueda

tener siguen siendo campo exclusivo del juez humano, brindando a la IA únicamente una labor silogística a través de los hechos que, como realidad judicial, ha fijado el juez humano.

No podemos desdeñar la importancia de una aplicación como la expuesta, dado que solo el hecho de que la IA entienda el lenguaje natural y pueda actuar en consecuencia ya es un avance impresionante, sin embargo, no le estamos, en ninguno de los dos casos expuestos, permitiendo que lleve a cabo una interpretación de la ley para el caso concreto, ni una valoración de la prueba.

El primero de los supuestos, la interpretación de una ley, no es una tarea que esté vedada para este tipo de IA, si bien no es el objeto del presente ensayo, quizás de otro, pero el segundo de los supuestos, la valoración de la prueba, es el más interesante.

GPT-4 es una IA multimodal, es decir, puede procesar y generar múltiples tipos de datos, como texto e imágenes, en el caso de GPT-4 puede recibir texto e imágenes y generar texto a partir de las mismas. Ello significa que GPT-4 podría apreciar una imagen y resolver en consecuencia. Un paso más allá sería la posibilidad de ver vídeos, escuchar audios y emitir texto. Esto nos llevaría a la posibilidad de aplicar técnicas de credibilidad en las pruebas personales (ReGIM-Lab Lie Detection DataBase –Nasri *et al.*, 2016–) y análisis de la congruencia o veracidad de documentos no digitalizados.

Antes de que se pueda considerar la implementación de jueces robot, es esencial establecer un marco regulatorio claro que aborde los desafíos éticos y técnicos⁶. Este marco debe garantizar la transparencia, la rendición de cuentas y la protección contra el sesgo y la discriminación.

En definitiva, la evolución en el ámbito del Deep Learning está ofreciendo una serie de herramientas que más tarde o más temprano forzarán a que la sociedad tome una decisión sobre qué tipo de justicia quiere y por quién quiere ser juzgada.

8. Código fuente

El código fuente de la aplicación está disponible en: <https://github.com/xvi82/juezesrobot>

Su uso requiere tener una API Key de OpenAI.

⁶ La *AI Act* es una propuesta de regulación presentada por la Comisión Europea en abril de 2021 para establecer reglas sobre el desarrollo y uso de la IA en la Unión Europea. La *AI Act* hace referencia a la aplicación de la IA en el ámbito de la justicia. La propuesta clasifica ciertas aplicaciones de IA en el ámbito judicial como de «alto riesgo» debido a las posibles implicaciones en los derechos fundamentales de las personas. Por lo tanto, estas aplicaciones estarían sujetas a requisitos más estrictos en términos de transparencia, responsabilidad y garantías para los usuarios.

Referencias bibliográficas

- Almeida, G. F. C. F., Nunes, J. L., Engelmann, N., Wiegmann, A. y De Araújo, M. (2023). *Exploring the psychology of GPT-4's Moral and Legal Reasoning*. <https://arxiv.org/abs/2308.01264v1>
- Ashley, K. D. (2017). *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*. Cambridge University Press. <https://doi.org/10.1017/9781316761380>
- Barocas, S. y Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671-732. <https://www.jstor.org/stable/24758720>
- Branting, L. K. (2017). Data-centric and logic-based models for automated legal problem solving. *Artificial Intelligence and Law*, 25, 5-27. <https://link.springer.com/article/10.1007/s10506-017-9193-x>
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C.,... Amodi, D. (2020). *Language models are few-shot learners*. <https://doi.org/10.48550/arXiv.2005.14165>
- Brownsword, R. (2008). *Rights, Regulation, and the Technological Revolution*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199276806.001.0001>
- Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. C. y Zhang, Y. (2023). *Sparks of Artificial General Intelligence: Early experiments with GPT-4*. <https://doi.org/10.48550/arXiv.2303.12712>
- Chouldechova, A. y Roth, A. (2018). *The frontiers of fairness in machine learning*. <https://doi.org/10.48550/arXiv.1810.08810>
- Dressel, J. y Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, 4(1). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5777393/>
- Han, S. J., Ransom, K., Perfors, A. y Kemp, C. (2023). *Inductive reasoning in humans and large language models*. <https://doi.org/10.48550/arXiv.2306.06548>
- Mutlu, B. y Forlizzi, J. (2008). *Robots in organizations: The role of workflow, social, and environmental factors in human-robot interaction*. HRI '08 Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction, 287-294. <https://dl.acm.org/doi/10.1145/1349822.1349860>
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S. y Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2). <https://doi.org/10.1177/205395171667967>
- Moor, J. H. (2006). The nature, importance, and difficulty of machine ethics. *IEEE intelligent systems*, 21, 18-21. <https://philpapers.org/rec/MOOTNI>
- Nasri, H., Ouarda, W. y Alimi, A. M. (2016). ReLiDSS: Novel lie detection system from speech signal. *IEEE/ACS 2016, 13.ª Conferencia Internacional de Sistemas y Aplicaciones Informáticas (AICCSA)*, 1-8. <https://doi.org/10.1109/AICCSA.2016.7945789>
- Nori, H., King, N., McKinney, S. M., Carignan, D. y Horvitz, E. (2023). *Capabilities of GPT-4 on Medical Challenge Problems*. <https://doi.org/10.48550/arXiv.2303.13375>
- Oswald, M., Grace, J., Urwin, S. y Barnes, G. C. (2018). Algorithmic risk assessment policing models: lessons from the Durham HART model and "Experimental" proportionality. *Information & Communications Technology*

- Law*, 27(2), 223-250. <https://doi.org/10.1080/13600834.2018.1458455>
- Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
- Peng, B., Li, C., He, P., Galley, M. y Gao, J. (2023). *Instruction Tuning with GPT-4*. <https://doi.org/10.48550/arXiv.2304.03277>
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D. y Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8), 9. https://d4mucfpksywv.cloudfront.net/better-language-models/language_models_are_unsupervised_multitask_learners.pdf
- Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J. F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., Jackson, M. O., Jennings, N. R., Kamar, E., Kloumann, I. M., Larochelle, H., Lazer, D., McElreath, R., Mislove, A., Parkes, D. C., Pentland, A. S. ... Wellman, M. (2019). Machine behaviour. *Nature*, 568, 477-486. <https://doi.org/10.1038/s41586-019-1138-y>
- Remus, D. y Levy, F. S. (2016). Can Robots Be Lawyers? Computers, Lawyers, and the Practice of Law. *Geo. J. Legal Ethics*, 30, 501-558. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2701092
- Russell, S. J. y Norvig, P. (2010). *Artificial Intelligence: A Modern Approach*. Prentice Hall. https://people.engr.tamu.edu/guni/csce421/files/AI_Russell_Norvig.pdf
- Sang-Hun, C. (15 de marzo de 2016). *Google's Computer Program Beats Lee Se-dol in Go Tournament*. <https://www.nytimes.com/2016/03/16/world/asia/korea-alphago-vs-lee-sedol-go.html>
- Schraudolph, N. N., Dayan, P. y Sejnowski, T. J. (1994). *Temporal Difference Learning of Position Evaluation in the Game of Go*. https://www.researchgate.net/publication/2301607_Temporal_Difference_Learning_of_Position_Evaluation_in_the_Game_of_Go
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Kalchbrenner, N., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T. y Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529, 484-489. <https://doi.org/10.1038/nature16961>
- Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Panadero, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifré, L., Van Den Driessche, G., Graepel, T. y Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature*, 550, 354-359. <https://doi.org/10.1038/nature24270>
- Surden, H. (2014). Machine learning and law. *Washington Law Review*, 89(1), 87-115. <https://digitalcommons.law.uw.edu/wlr/vol89/iss1/5/>
- Susskind, R. (2019). *Online Courts and the Future of Justice*. Oxford University Press. <https://global.oup.com/academic/product/online-courts-and-the-future-of-justice-9780192849304?cc=us&lang=en&>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. y Polosukhin, I. (2017). *Attention is all you need*. *Advances in neural information processing systems*, 30. https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html
- Webb, T., Holyoak, K. J. y Lu, H. (2022). *Emergent Analogical Reasoning in Large Language Models*. <https://doi.org/10.48550/arXiv.2212.09196>



White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J. y Schmidt, D. C. (2023). *A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT*. <https://doi.org/10.48550/arXiv.2302.11382>

Zarsky, T. Z. (2013). Transparent predictions. *University of Illinois Law Review*, 4, 1.503-1.570. <https://www.illinoislawreview.org/wp-content/ilr-content/articles/2013/4/Zarsky.pdf>

Javier Ercilla García. Magistrado especialista en jurisdicción social. Cuenta con una sólida formación jurídica y técnica. Destaca por sus publicaciones y ponencias sobre la incidencia de las nuevas tecnologías, especialmente la inteligencia artificial y la robótica en el ámbito laboral y la Administración de justicia. Ha sido galardonado con el Premio a la Calidad de la Justicia 2020 en la modalidad «Justicia más eficaz» y es autor de diversos proyectos de automatización y gestión de sentencias utilizando lenguajes de programación como Python. <https://orcid.org/0009-0006-5930-2574>