



APM 3.9.

Publicación digital. - Asociación Profesional de
la Magistratura

JAVIER ERCILLA GARCÍA

CANARIAS

INTELIGENCIA ARTIFICIAL Y REVISIÓN FÁCTICA EN LA JURISDICCIÓN SOCIAL: INTEGRANDO GPT-4 EN LA RESOLUCIÓN DE LOS RECURSOS DE SUPPLICACIÓN

ARTIFICIAL INTELLIGENCE AND FACTUAL REVISION IN SOCIAL
JURISDICTION: INTEGRATING GPT-4 IN THE RESOLUTION OF APPEALS

Resumen

Este artículo presenta un programa desarrollado en Tkinter utilizando el motor de IA GPT-4, de OpenAI, para resolver revisiones fácticas en recursos de suplicación en la jurisdicción social. Se describe el diseño de la interfaz gráfica y la implementación de GPT-4 con un contexto jurídico específico. El programa permite la corrección manual y el aprendizaje de nuevas reglas a través de explicaciones. Se analizan las repercusiones actuales, como el impacto en la eficiencia de las resoluciones judiciales y el cambio en el rol de los profesionales del derecho. Finalmente, se exploran las posibilidades futuras del programa, incluyendo mejoras en la precisión y capacidad de aprendizaje de GPT-4, expansión a otras áreas del derecho y la integración de tecnologías complementarias.

Palabras clave

Revisión Fáctica. Inteligencia Artificial. Generación de Texto. Recursos de Suplicación. Jurisdicción Social. Contexto Jurídico. Corrección Manual. Aprendizaje. Eficiencia. Resoluciones Judiciales. Profesionales del Derecho. Inteligencia Artificial. Evolución. Precisión. Capacidad de Aprendizaje. Expansión. Tecnologías Complementarias. Procesamiento del Lenguaje Natural. Impacto. Tkinter. GPT-4. Python. Código Fuente.

Abstract

This article presents a Tkinter-based program utilizing the GPT-4 engine, from OpenAI, to address factual revisions in appeals within the social jurisdiction. The graphical interface design and GPT-4 implementation with a specific legal context are described. The program allows manual correction and rule learning through explanations. Current implications are analyzed, such as the impact on judicial resolutions' efficiency and the shift in legal professionals' roles. Finally, the program's future possibilities are explored, including improvements in GPT-4's accuracy and learning capacity, expansion into other areas of law, and the integration of complementary technologies.

Keywords

Factual Revision. Artificial Intelligence. Text Generation. Appeals. Social Jurisdiction. Legal Context. Manual Correction. Learning. Efficiency. Judicial Resolutions. Legal Professionals. Artificial Intelligence. Evolution. Accuracy. Learning Capacity. Expansion. Complementary Technologies. Natural Language Processing. Impact. Tkinter. GPT-4. Python. Source Code.

Sumario

1. Introducción. 1.1. Contexto y propósito del programa. 1.2. Antecedentes y evolución de GPT-4. 2. Desarrollo del programa en Tkinter. 2.1. Diseño de la interfaz gráfica. 2.2. Implementación de GPT-4 y el context-length. 2.2.1 Prompt engineering. 2.2.2 Prompting. 2.2.3 El prompt de la revisión fáctica. 2.3. Proceso de corrección y aprendizaje. 2.4. Integración de la API de Open AI. 2.5. Pruebas y evaluación del programa. 2.6. Limitaciones. 2.7. Consideraciones éticas y de privacidad. 3. Repercusiones. 3.1. Impacto en la eficiencia y calidad de las resoluciones judiciales. 3.2. Cambios en el rol de los profesionales del derecho. 3.3. Desventajas de la implementación de la IA en la revisión fáctica. 4. Posibilidades futuras y evolución del programa. 4.1. Mejoras en la precisión y capacidad de aprendizaje de GPT-4. 4.2. Expansión. 4.3. Integración de tecnologías complementarias y avances en el procesamiento del lenguaje natural con bases de datos jurídicas. 5. Conclusiones. 6. Código fuente. 7. Referencias.

1. Introducción

La Jurisdicción Social se configura como una jurisdicción de instancia única, de tal manera que los recursos que puedan interponerse frente a las sentencias dictadas por los Juzgados de lo Social son extraordinarios, esto significa que únicamente permiten revisar alguna faceta del enjuiciamiento (solo la aplicación e interpretación del derecho), sometiéndose a un elenco tasado de motivos de impugnación fuera de los cuales el recurso deviene improcedente.

El recurso que puede interponerse frente a las sentencias dictadas por los Juzgados de lo Social es el recurso de suplicación. La Ley de 17 de octubre de 1940 fue la primera en mencionar el recurso de suplicación, que se presentaba contra las sentencias emitidas por las Magistraturas de Trabajo y era resuelto por el Tribunal Central de Trabajo. En 1989, con la creación de los Tribunales Superiores de Justicia y la desaparición del Tribunal Central de Trabajo, la Ley de Procedimiento Laboral de 1990 se adaptó al cambio, y desde entonces, los recursos de suplicación son resueltos por los Tribunales Superiores de Justicia.

El artículo 190 de la Ley Reguladora de la Jurisdicción Social (LRJS) establece que las Salas de lo Social de los Tribunales Superiores de Justicia conocerán de los recursos de suplicación interpuestos contra resoluciones dictadas por los Juzgados de lo Social, así como contra autos y sentencias dictadas por Jueces de lo Mercantil que afecten al derecho laboral.

El recurso de suplicación tiene una naturaleza extraordinaria, casi casacional y de objeto limitado, lo que implica: 1) Limitación de las resoluciones recurribles; 2) Articulación solo a través de motivos tasados legalmente; 3) Cumplimiento estricto de requisitos formales para evitar desestimación del recurso; 4) Imposibilidad del Tribunal ad quem de realizar nueva valoración de la prueba o resolver sobre cuestiones distintas a las planteadas; y 5) Aplicación de la regulación de casación civil, no de apelación, debido a la naturaleza extraordinaria del recurso.

Dentro de los motivos tasados por los que puede impugnarse la sentencia de instancia tenemos la revisión fáctica. A saber, el art. 193.b) LRJS dispone lo siguiente:

"El recurso de suplicación tendrá por objeto:

[...]

b) Revisar los hechos declarados probados, a la vista de las pruebas documentales y periciales practicadas."

La revisión fáctica parte de diversas premisas y requisitos, resumidamente, 1) Señalar con precisión el hecho considerado erróneo; 2) Ofrecer un texto alternativo para corregir el error; 3) Citar documentos o pericias que demuestren la equivocación, sin menciones genéricas; 4) Los documentos deben mostrar el error de forma clara,

directa y patente, sin conjeturas ni argumentaciones; 5) La revisión debe ser trascendente a la sentencia, con efectos modificadores; 6) No se debe tratar de una nueva valoración global de la prueba incorporada al proceso.

Ahora bien, hay una serie de elementos, fundamentalmente técnicos, que, de la simple lectura del recurso, sin necesidad de entrar a un análisis más pormenorizado de la prueba (documental o pericial), pueden facilitar la denegación de la revisión fáctica. Son los supuestos de revisiones basadas en testificales, interrogatorios, revisiones de hechos conformes o notorios, hechos negativos, predeterminantes del fallo, indirectos, etc.

Estas cuestiones técnicas en ocasiones son ciertamente mecánicas, sin embargo, ello no impide que sea necesaria la lectura del recurso, su estudio y la redacción de la respuesta adecuada, esto es, la comprensión de lo que se pide y cómo se pide. Todas estas tareas requieren una labor de cognición, si bien no profunda, si concreta. Estas tareas no eran susceptibles de automatización hasta que se desarrollaron los grandes modelos de lenguaje natural (en adelante LLM¹).

El desarrollo de tecnologías de inteligencia artificial (en adelante IA) y procesamiento del lenguaje natural (en adelante PLN) ha permitido la creación de herramientas que pueden asistir en la realización de tareas jurídicas, incluida la revisión fáctica. En este contexto, este artículo presenta un programa desarrollado en Tkinter (un binding de la biblioteca gráfica Tcl/Tk para el lenguaje de programación Python) que utiliza la arquitectura GPT-4 de OpenAI para abordar la revisión fáctica en los recursos de suplicación. El programa integra la capacidad de GPT-4 para procesar y aprender de grandes volúmenes de información, ofreciendo una herramienta útil y eficiente para los Magistrados a la hora de resolver los motivos de revisión fáctica del recurso de suplicación.

1.1.Contexto y propósito del programa

El programa desarrollado en Tkinter no deja de ser una prueba de concepto, un modelo conceptual que expone cuál es el estado del arte actual en lo que se refiere a combinación entre IA y Derecho, y hasta donde podríamos llegar. El programa tendría como objetivo plausible asistir a magistrados en la revisión fáctica de los recursos de suplicación. Sería un asistente de IA, esto es, una tecnología habilitadora, y no de sustitución. Esta herramienta ofrecería al Magistrado un dictamen - estimatorio o desestimatorio - sobre la revisión fáctica propuesta en el recurso.

¹ Large Language Model, un tipo de modelo de aprendizaje automático que puede realizar una variedad de tareas de procesamiento de lenguaje natural (NLP), incluida la generación y clasificación de texto, responder preguntas de manera conversacional y traducir texto de un idioma a otro.

La herramienta se basa en la capacidad de GPT-4 para comprender y generar texto en lenguaje natural, así como para razonar y aprender de la información proporcionada. El programa cuenta con dos cuadros de texto en su interfaz gráfica: uno para introducir el motivo de revisión fáctica propuesto por el recurrente y otro para que GPT-4 proporcione una respuesta fundamentada sobre si se estima o desestima la revisión.

La utilización de GPT-4 en este contexto permite aprovechar sus mejoras en cuanto a capacidad de aprendizaje y procesamiento de información, en comparación con versiones anteriores como GPT-3.5. En concreto, la versión actual de GPT-4 permite trabajar con un context-length (ventana de contexto) de hasta 8,192 tokens, lo que equivale aproximadamente a 12 páginas de texto. Esta característica posibilita proporcionar un contexto amplio y detallado a GPT-4, incluyendo información sobre la jurisprudencia y la doctrina aplicable al caso, lo que facilita que la herramienta genere respuestas coherentes y adecuadas al marco legal vigente.

Además, el programa incluye una opción para que el usuario pueda corregir manualmente las respuestas generadas por GPT-4 en caso de errores o imprecisiones. Dicha corrección se utiliza para enseñar al modelo cómo abordar correctamente situaciones similares en el futuro, mediante la creación de reglas que se incorporan al contexto proporcionado. De esta manera, el programa se vuelve más preciso y eficiente a medida que se utiliza y se enfrenta a diferentes casos y situaciones.

1.2. Antecedentes y evolución de GPT-4

GPT-4 es una arquitectura de inteligencia artificial desarrollada por OpenAI, que forma parte de la serie de modelos de procesamiento del lenguaje natural conocidos como "Generative Pre-trained Transformers" (GPT). Estos modelos se basan en la técnica de "transformers", introducida por Vaswani et al. (2017)², que ha revolucionado el campo del procesamiento del lenguaje natural en los últimos años. Los transformers permiten el procesamiento paralelo de secuencias de texto y la atención a largas dependencias, lo que resulta en un mejor rendimiento en diversas tareas de PLN.

La evolución de los modelos GPT ha sido notable desde su primera versión, con mejoras significativas en términos de capacidad, rendimiento y aplicabilidad en una amplia gama de tareas. GPT-2, lanzado en 2019, mostró un avance considerable en la generación de texto coherente y relevante, aunque su uso estaba limitado por la restricción en la cantidad de tokens que podía procesar. GPT-3, lanzado en 2020, amplió aún más las capacidades de los modelos anteriores, permitiendo un mejor manejo de la información contextual y una mayor adaptabilidad a diversas tareas, incluyendo la generación de texto creativo, la

² Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. arXiv preprint arXiv:1706.03762. Disponible en: <https://arxiv.org/abs/1706.03762>

traducción automática y la resolución de problemas basados en lenguaje natural.

GPT-3.5, lanzado poco antes de GPT-4, ofrecía mejoras en el rendimiento y la eficiencia, pero mantenía algunas limitaciones en cuanto al context-length, lo que restringía la cantidad de información que el modelo podía utilizar para generar respuestas. GPT-3.5 permitía una ventana de contexto (context-length) de hasta 4.096 tokens, a saber, unas 3.000 palabras. Ello permitía dibujar a la IA un mínimo marco en torno al cual realizar las tareas que se le solicitasen. Con la introducción de GPT-4, se ha logrado superar estas limitaciones, permitiendo el procesamiento de hasta 8.192 tokens en un único contexto, a saber, unas 6.000 palabras. Pero es más, estos 8.192 tokens se ofrecen para la versión GPT-4 8k, habiendo una versión 32k, que permite un contexto de 32.768 tokens, ello serían unas 24.000 palabras, a saber, unas 50 páginas de texto. Esto representa un avance crucial para aplicaciones como la revisión fáctica en el ámbito jurídico, donde es necesario proporcionar información detallada y compleja para obtener resultados precisos y coherentes con el marco legal vigente.

Nos encontramos, por tanto, ante una IA que tiene la posibilidad de llevar a cabo tareas cognitivas que hasta ahora estaban limitadas a los seres humanos. Esta IA, antes, sólo podía ser 'enseñada' a través de instrucciones de no más de 3 o 6 páginas, sin embargo, ahora, con GPT-4 podemos enseñar a dicha IA con hasta 50 páginas de contexto. Imaginemos cómo actuaría un Juzgador humano tras la lectura de un artículo doctrinal de 50 páginas sobre la revisión fáctica y analicemos cómo podría actuar una IA en ese mismo contexto.

La arquitectura GPT-4, lanzada el 15 de marzo de 2023, está demostrando ser eficaz en una amplia gama de tareas de procesamiento del lenguaje natural, incluyendo la generación de texto, la respuesta a preguntas, la clasificación de texto, la traducción automática y la extracción de información, entre otras. Su capacidad para adaptarse a diferentes dominios y aprender de grandes volúmenes de datos lo convierte en una herramienta valiosa para abordar desafíos específicos, como la revisión fáctica en el ámbito jurídico.

El desarrollo del programa en Tkinter que utiliza GPT-4 para asistir en la revisión fáctica de los recursos de suplicación es un ejemplo concreto de cómo la inteligencia artificial y el procesamiento del lenguaje natural pueden contribuir a mejorar la eficiencia y la calidad de los procesos legales. Al proporcionar un contexto detallado y riguroso a GPT-4, el programa permite generar respuestas fundamentadas y coherentes con la jurisprudencia y la doctrina aplicables. Además, la capacidad de corregir y enseñar al modelo a partir de errores o imprecisiones garantiza que la herramienta se vuelva cada vez más precisa y confiable a medida que se utiliza en diferentes casos y situaciones.

El contexto en las IA generativas de texto es esencial. En los modelos GPT-3.5 y GPT-4, este contexto se enmarca en su configuración como chatbot. En los modelos de GPT-3, el contexto o la manera de responder requería de un precio 'ajuste fino' o fine tuning, que no será objeto del presente artículo. En este sentido, como señala Jacob Stern en su artículo en The Atlantic, GPT-4 tiene una 'memoria de pez'³. A pesar de los avances en los modelos de lenguaje basados en IA, como GPT-4, estos todavía presentan problemas de memoria, sesgo y falta de sentido común. Los modelos de lenguaje actuales tienen una memoria deficiente, lo que limita su capacidad para realizar tareas complejas. Aunque GPT-4 puede realizar tareas como codificar sitios web o redactar correos electrónicos, su memoria a corto plazo es limitada. Esto se debe a que solo puede tener en cuenta una cantidad limitada de texto, conocida como ventana de contexto.

La ventana de contexto de GPT-4 es de unas 12 páginas (6.000 palabras), y por ende es lo suficientemente larga como para mantener conversaciones similares a las humanas, pero aún limitada para tareas más complicadas como resumir libros completos o revisar proyectos de programación. Ahora bien, esta ventana, ya sea la de GPT-4 8k (12 páginas) o la de GPT-4 32k (50 páginas), nos sirve para proyectos como el presente, a saber, resolver problemas sencillos, esencialmente técnicos, en el recurso de suplicación.

La falta de memoria a largo plazo es otro problema de GPT-4. Incluso si la ventana de contexto se ampliara significativamente, cada vez que se inicie una nueva conversación con GPT-4, comenzaría desde cero. Esta limitación, si bien no es resuelta, si es parcheada en el presente proyecto, mediante la actualización de la ventana de contexto con la retroalimentación del usuario ante respuestas erróneas de GPT-4, como analizaremos a lo largo del artículo.

2. Desarrollo del programa en Tkinter

Tkinter es una biblioteca de Python que permite la creación de interfaces gráficas de usuario (GUI) de manera sencilla y efectiva. El desarrollo del programa para la revisión fáctica de recursos de suplicación en el ámbito jurídico se basa en esta biblioteca, aprovechando sus características y la capacidad del modelo GPT-4 para generar respuestas fundamentadas y coherentes con el contexto proporcionado.

2.1. Diseño de la interfaz gráfica

La interfaz gráfica del programa en Tkinter se diseñó pensando en la simplicidad y funcionalidad para el usuario. Consta de dos cuadros de texto y un conjunto de botones de acción. A continuación, se describen los principales elementos de la interfaz:

³ Stern, J. (2023) GPT-4 Has the Memory of a Goldfish. The Atlantic. Disponible en: <https://www.theatlantic.com/technology/archive/2023/03/gpt-4-has-memory-context-window/673426/>

- Cuadro de texto de entrada: Este cuadro de texto permite al usuario introducir el motivo de revisión fáctica propuesto por el recurrente. Se diseñó para admitir la entrada de texto en formato de párrafo, lo que facilita la introducción de argumentos y detalles relevantes del caso.
- Cuadro de texto de salida: En este cuadro de texto, GPT-4 proporciona la respuesta fundamentada en función del contexto y la información proporcionada en el cuadro de texto de entrada. La respuesta incluye un análisis y razonamiento sobre si se estima o desestima la revisión fáctica, basándose en la jurisprudencia y doctrina aplicables.
- Botones de acción: La interfaz gráfica incluye botones de acción que permiten al usuario interactuar con el programa. Estos botones incluyen:

a) "Generar respuesta": Al hacer clic en este botón, con el título de 'Valorar', se activa el proceso de generación de respuesta por parte de GPT-4, tomando en cuenta la información proporcionada en el cuadro de texto de entrada.

b) "Corregir respuesta": Este botón permite al usuario corregir manualmente la respuesta generada por GPT-4 en caso de errores o imprecisiones. Al hacer clic en este botón, se abre una ventana adicional en la que el usuario puede ingresar la respuesta correcta y una explicación sobre el error cometido por GPT-4.

c) "Información": Al hacer clic en este botón, mostrará información bajo los cuadros de texto en los que se explicará cada uno de ellos.

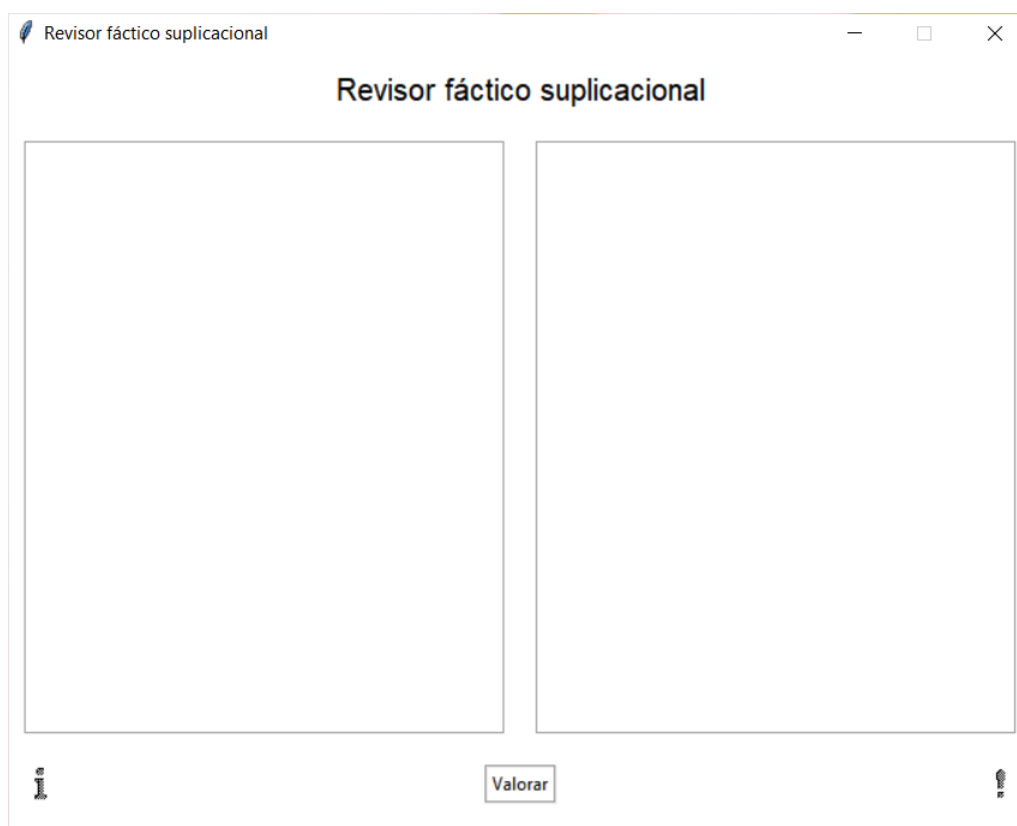


Imagen 1: La interfaz gráfica del programa en Tkinter muestra los cuadros de texto de entrada y salida, así como los botones de acción para generar respuesta, corregir respuesta (;) e información (i).

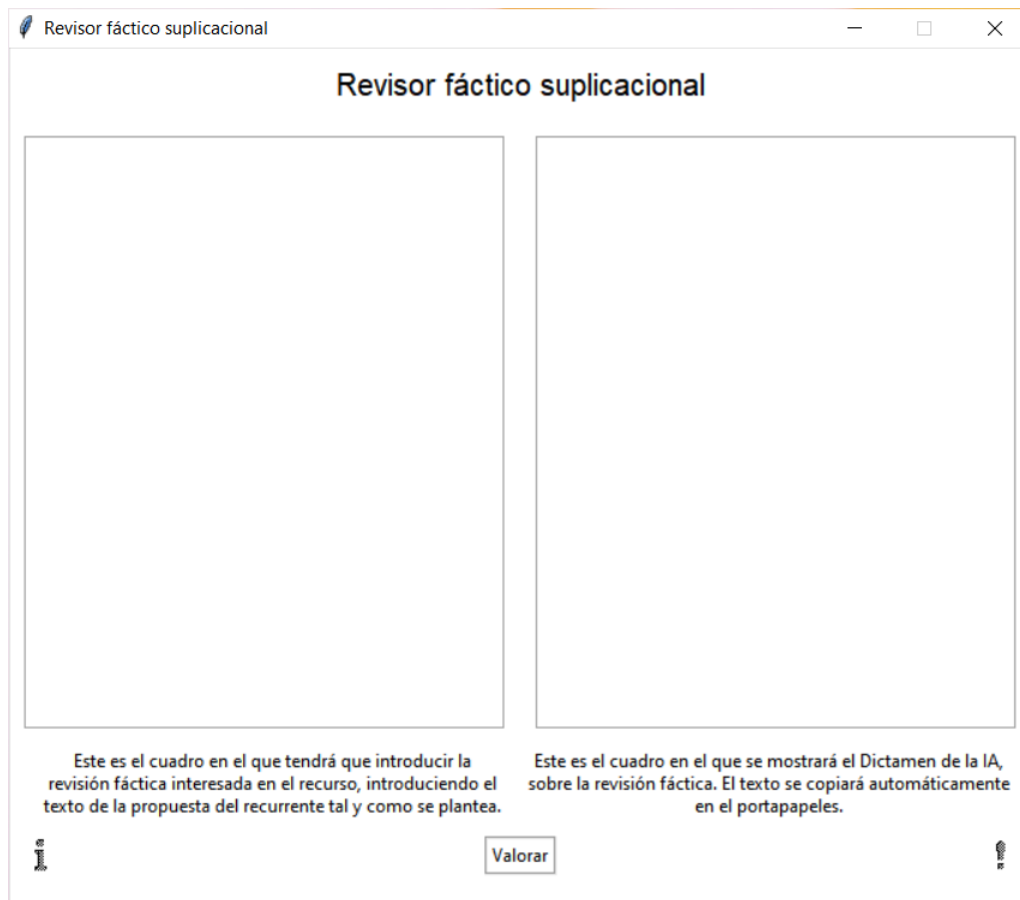


Imagen 2: La interfaz gráfica del programa en Tkinter tras pulsar el botón 'i' de información.

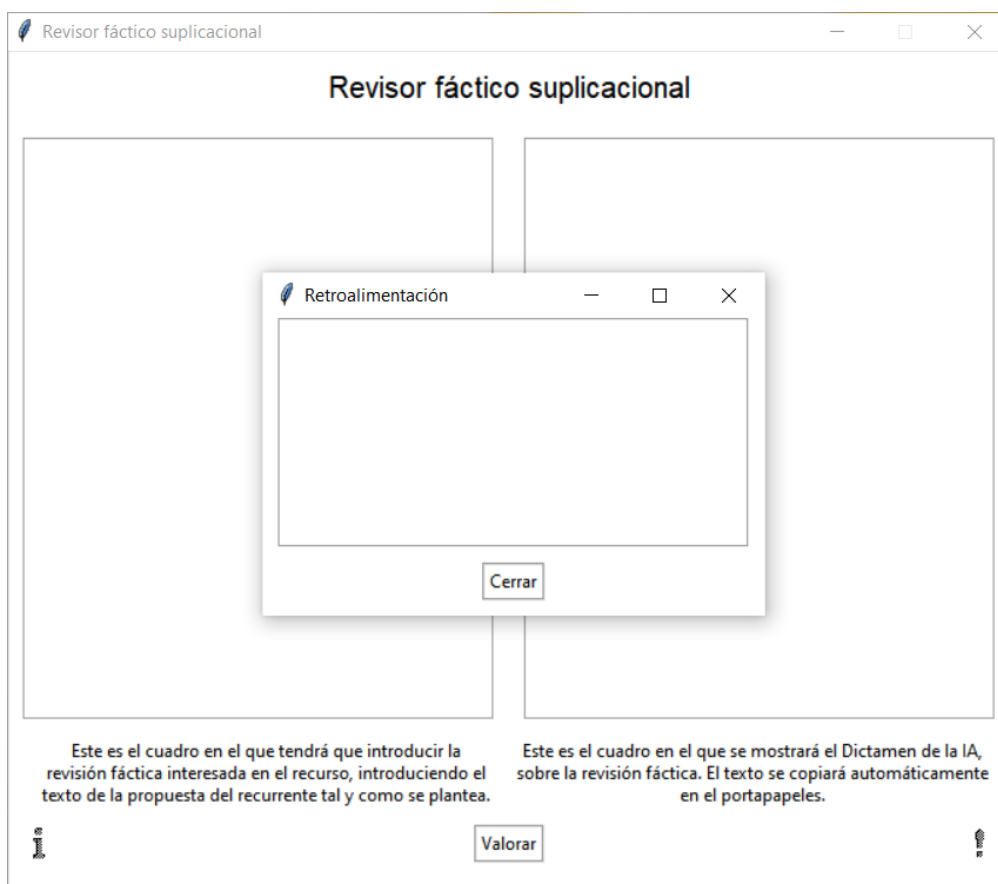


Imagen 3: La interfaz gráfica del programa en Tkinter tras pulsar el botón '!' de corregir respuesta. Aparece un nueva ventana con un cuadro de texto. En él el usuario introducirá la respuesta correcta.

2.2. Implementación de GPT-4 y el context-length

Una de las características clave de GPT-4 es su capacidad para manejar una ventana de contexto (context-length) más amplia en comparación con versiones anteriores, como GPT-3.5. Esto permite proporcionar un contexto detallado y riguroso al modelo, lo que resulta en respuestas más precisas y coherentes con el marco legal vigente. GPT-4 admite una ventana de contexto de hasta 8,192 tokens, lo que equivale aproximadamente a 12 páginas de texto.

Para el desarrollo del programa en Tkinter, se utilizó esta capacidad para proporcionar un contexto de aproximadamente 6 páginas a GPT-4. Este contexto incluye información relevante sobre cómo se resuelve la revisión fáctica en el ámbito jurídico, así como citas de jurisprudencia y doctrina aplicables. Al proporcionar un contexto detallado y específico, se logra que GPT-4 genere respuestas que estén en consonancia con la jurisprudencia del Tribunal Supremo en torno a la revisión del relato fáctico.

2.2.1. Prompt engineering

Para llevar a cabo una ventana de contexto adecuada hay que saber la forma en la que hay que dirigirse a la IA, a GPT-4. La forma en la que el usuario interactúa con GPT-4 es a través de 'prompts' o

'entradas'. Para que cualquier modelo de GPT realice una tarea precisa y detallada sin sobrepasar la encomienda, es necesario formular la 'entrada' o 'prompt' siguiendo ciertos parámetros específicos, de manera que se le indique a la IA qué es exactamente lo que se busca obtener. Estos parámetros y la creación de los mismos se conocen como 'prompt engineering' y consisten en la definición en la entrada de las tareas concretas y el modo en que la IA generará el texto.

La estructura de toda 'entrada', para obtener resultados óptimos, sería la siguiente:

<p>Lorem ipsum dolor sit amet, consectetur adipiscing elit. Proin efficitur blandit ex sit amet venenatis. Mauris maximus dapibus leo, eget iaculis risus imperdiet eget. Nulla nec turpis elit. Pellentesque lobortis consectetur auctor. Morbi urna dolor, placerat vel feugiat a, mollis nec nisi. Nullam id tempor metus, vel congue ante. Aenean rutrum justo turpis, eget placerat lectus euismod eu.</p>	<p>Sujeto a simular</p>
<p>Pellentesque sit amet ipsum sed erat tincidunt facilisis non non erat. Aliquam lacinia nisi nulla, ut tristique justo interdum vel. Quisque vestibulum risus neque, non blandit massa scelerisque nec. Praesent id luctus ante. Nullam augue felis, mollis rutrum posuere at, convallis eu ex. Morbi at tempor justo. In hac habitasse platea dictumst. Fusce eleifend enim tellus, a consectetur velit pretium ac. Phasellus eget justo dolor.</p>	<p>Tarea</p>
<p>Pellentesque sit amet ipsum sed erat tincidunt facilisis non non erat. Aliquam lacinia nisi nulla, ut tristique justo interdum vel. Quisque vestibulum risus neque, non blandit massa scelerisque nec. Praesent id luctus ante. Nullam augue felis, mollis rutrum posuere at, convallis eu ex. Morbi at tempor justo. In hac habitasse platea dictumst. Fusce eleifend enim tellus, a consectetur velit pretium ac. Phasellus eget justo dolor.</p>	<p>Pasos para completar la tarea</p>
<p>Pellentesque sit amet ipsum sed erat tincidunt facilisis non non erat. Aliquam lacinia nisi nulla, ut tristique justo interdum vel. Quisque vestibulum risus neque, non blandit massa scelerisque nec. Praesent id luctus ante. Nullam augue felis, mollis rutrum posuere at, convallis eu ex. Morbi at tempor justo. In hac habitasse platea dictumst. Fusce eleifend enim tellus, a consectetur velit pretium ac. Phasellus eget justo dolor.</p>	<p>Contexto / Restricciones</p>
<p>Pellentesque sit amet ipsum sed erat tincidunt facilisis non non erat. Aliquam lacinia nisi nulla, ut tristique justo interdum vel. Quisque vestibulum risus neque, non blandit massa scelerisque nec. Praesent id luctus ante. Nullam augue felis, mollis rutrum posuere at, convallis eu ex. Morbi at tempor justo. In hac habitasse platea dictumst. Fusce eleifend enim tellus, a consectetur velit pretium ac. Phasellus eget justo dolor.</p>	<p>Objetivo</p>
<p>Pellentesque sit amet ipsum sed erat tincidunt facilisis non non erat. Aliquam lacinia nisi nulla, ut tristique justo interdum vel. Quisque vestibulum risus neque, non blandit massa scelerisque nec. Praesent id luctus ante. Nullam augue felis, mollis rutrum posuere at, convallis eu ex. Morbi at tempor justo. In hac habitasse platea dictumst. Fusce eleifend enim tellus, a consectetur velit pretium ac. Phasellus eget justo dolor.</p>	<p>Formato de Salida</p>

En esta tabla es muestra una a la izquierda una 'entrada' o 'prompt' ficticio (lorem ipsum) y a la derecha las partes en que debería articularse dicho 'prompt' para tener éxito en nuestra prueba.

En el contexto que creamos para que GPT-4 pueda resolver la revisión fáctica de un recurso de suplicación hemos de empezar por determinar cuál es el rol que la IA asume. Esto es, hemos de concretar el sujeto a simular, en nuestro caso, el 'prompt' comenzaría así:

"Eres un Juez, en un Tribunal [...]"

Podríamos técnicamente criticar que en un Tribunal habría Magistrados y no Jueces, pero hemos de tener en cuenta que la información que suministremos en nuestro 'prompt' ha de ser sencilla, univoca, y sin posibilidad de interpretaciones alternativas, por lo que la palabra

Juez resulta la más adecuada para que la IA sepa qué rol asumir.

El siguiente paso es introducir la tarea a realizar:

"debes resolver un recurso de suplicación frente a una Sentencia"

La parte más importante del 'prompt' es la de los pasos para completar la tarea, si bien debido a su complejidad la abordaremos más adelante.

A continuación, señalaremos el contexto y/o restricciones, a saber:

"En el recurso te piden la revisión de los Hechos Probados de la Sentencia"

"Las Sentencias tienen Hechos Probados, Fundamentos de Derecho y el Fallo."

"La revisión de los Hechos Probados se articula legalmente a través del art. 193.a) de la Ley Reguladora de la Jurisdicción Social (en adelante LRJS)."

"El apartado de la sentencia denominado "Hechos Probados" fija la versión judicial de los hechos litigiosos"

Estas son algunas de las frases de contexto y/o restricciones que habrán de darse a GPT-4 en la ventana de contexto antes de que comience a realizar las tareas que queremos, a saber, resolver la revisión fáctica.

El objetivo puede confundirse en ocasiones con la tarea, a saber:

"debes resolver un recurso de suplicación frente a una Sentencia"

Por último, en el formato de salida, pueden utilizarse 'prompts' como que el formato sea un esquema, una lista, una tabla, etc. En nuestro caso señalaremos:

"utiliza lenguaje jurídico, emplea una estructura IRAC (no lo dividas en apartados), el texto debe ser argumentativo"


Le estamos diciendo a la IA cuál es el formato que queremos, a saber, el tono del texto (jurídico) la forma de desarrollo (argumentativa) y hasta la estructura de resolución (IRAC⁴).

⁴ Es una técnica de argumentación legal, acrónimo de las palabras Issue, Rule, Application, Conclusion, es una forma de razonamiento deductivo que se complementa con la argumentación por analogía.

2.2.2. Prompting

Volviendo a los pasos para completar la tarea, este es el punto esencial del 'prompt' y aquí diferenciaremos varios tipos de 'prompting', de los cuales dependerá cuán eficaz es la respuesta que obtendremos de GPT-4. Por una parte, tenemos el "Standard prompt"⁵, que consiste únicamente en una pregunta:

JE ¿Qué es el recurso de suplicación?

 El recurso de suplicación es un recurso que se presenta ante una instancia superior para que revise una sentencia emitida por un tribunal inferior. Es decir, es una herramienta legal que permite a una parte impugnar la decisión tomada por un juez o tribunal en un proceso judicial.

El recurso de suplicación se utiliza en muchos sistemas judiciales como una forma de revisión y control de las decisiones judiciales, y generalmente está disponible para las partes que no estén satisfechas con el resultado de un juicio.

En general, el recurso de suplicación se presenta ante un tribunal superior al que emitió la sentencia original, y debe contener los argumentos y fundamentos que justifiquen la impugnación. El tribunal superior revisará el recurso y podrá confirmar, modificar o anular la sentencia original, dependiendo de los argumentos presentados y de la evaluación del caso en cuestión.

Por otro lado, tenemos los 'Few shot standard prompts'⁶, que son simplemente 'prompts' estándar con ejemplos (exemplars) en ellos, estos ejemplos ayudan a la IA a tener un contexto sobre el formato y forma en que quiere que se realice la tarea subsiguiente:

JE

Q: ¿En qué fecha se publicó el Código Civil?

A: El Código Civil fue publicado el 24 de julio de 1889 en la Gaceta de Madrid.

Q: ¿En qué fecha se publicó el Código Penal de 1995?



El Código Penal de 1995 fue publicado en el Boletín Oficial del Estado (BOE) el 23 de noviembre de 1995.

Así mismo, otra forma de hacer 'prompting' sería la 'Chain of Thought'⁷, o cadena de pensamiento. Esta forma de 'prompting' es la que nos interesa en el presente proyecto, dado que fomenta que la IA exponga una explicación clara del razonamiento que le lleva a la decisión. Es decir, que explique por qué estima o desestima la revisión fáctica planteada, en definitiva, motivar. Lo que hace el 'prompt' es mostrar una tarea, resolverla y explicar el razonamiento lógico de dicha resolución, para que la IA contextualice la forma

⁵ Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., & Iwasawa, Y. (2022). Large Language Models are Zero-Shot Reasoners. Disponible en: <https://arxiv.org/abs/2205.11916>

⁶ Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2022). Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. ACM Computing Surveys. Disponible en: <https://doi.org/10.1145/3560815>

⁷ Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2022). Chain of Thought Prompting Elicits Reasoning in Large Language Models. Disponible en: <https://arxiv.org/abs/2201.11903>

en que han de resolverse los problemas:

En nuestro caso, cuando expliquemos a la IA cómo se resuelven las cuestiones técnicas de la revisión fáctica (hechos negativos, hechos predeterminantes del fallo, obstrucción negativa, etc.), habremos de mostrar ejemplos de cada una de esas cuestiones y cómo se han resuelto, de tal manera que la IA contextualice la explicación en una materialización concreta.

JE

¿Cuál es la forma más rápida de llegar a la universidad?

Option 1: Tomar un autobús de 20 minutos, luego un metro de 25 minutos y finalmente caminar 10 minutos.

Option 2: Tomar un autobús de 30 minutos, luego un metro de 15 minutos y finalmente caminar 5 minutos.

Option 1: tomará $20+25+10=55$ minutos.

Option 2: tomará $30+15+5=50$ minutos.

Dado que la opción 2 tarda 50 minutos y la opción 1 tarda 55 minutos, la opción 2 es más rápida para llegar a la universidad.

¿Cuál es la forma más rápida de llegar al juzgado durante la hora pico del tráfico?

Option 1: Tomar un autobús de 20 minutos, luego un patinete eléctrico durante 25 minutos, un metro de 15 minutos y finalmente caminar 10 minutos.

Option 2: Tomar un tren de cercanías de 30 minutos, luego caminar 5 minutos, tomar un metro de 20 minutos, un tranvía de 10 minutos y finalmente caminar 5 minutos.



Option 1: tomará $20+25+15+10=70$ minutos.

Option 2: tomará $30+5+20+10+5=70$ minutos.

Ambas opciones tardarán lo mismo, 70 minutos, debido al tráfico en hora pico. Por lo tanto, no hay opción más rápida para llegar al juzgado en este caso.

Otra forma de 'prompt' es la 'Zero Shot Chain of Thought'⁸, o cadena de pensamiento sin entrenamiento previo, consiste únicamente en indicar a la IA cómo hacer la tarea como la frase 'Pensemos paso a paso'. Resulta sorprendente cómo, las primeras versiones de GPT podían fallar en problemas matemáticos simples, pero si se le añadía al 'prompt' la frase 'Pensemos paso a paso', lo resolvían correctamente. Con los modelos actuales de GPT, en concreto con GPT-4, esta técnica se ha superado, sin que la utilización de la misma de resultados mejores, únicamente se ha apreciado que los resultados son más extensos. En nuestro caso si utilizaremos la expresión 'Pensemos paso a paso' al objeto de obtener una redacción más extensa y una argumentación más prolija de la decisión tomada.

La última técnica de 'prompting' a tener en cuenta a los efectos de la ventana de contexto que hemos de crear para nuestro programa de revisión fáctica es la 'Generated Knowledge'⁹ o enfoque de conocimiento generado, que consiste en generar conocimiento a partir de un modelo de lenguaje, y luego proporcionar dicho conocimiento como aporte adicional al responder una pregunta. En el programa desarrollado el 'Generated Knowledge' será doble, por una parte, por toda la información que de sentencias y doctrina se ha utilizado para enseñar a la IA cómo resolver la revisión fáctica, y, por otra parte, por el conocimiento que se genera al introducir correcciones por parte del usuario, correcciones que generan conocimiento a través del propio GPT-4 y que este tendrá en

⁸ Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., & Iwasawa, Y. (2022). Large Language Models are Zero-Shot Reasoners. Disponible en: <https://arxiv.org/abs/2205.11916>

⁹ Liu, J., Liu, A., Lu, X., Welleck, S., West, P., Bras, R. L., Choi, Y., & Hajishirzi, H. (2021). Generated Knowledge Prompting for Commonsense Reasoning. Disponible en: <https://arxiv.org/abs/2110.08387>

cuenta para los futuros recursos.

2.2.3. El prompt de la revisión fáctica

Una vez explicado todo lo que antecede, vamos a analizar someramente el 'prompt' o entrada de la ventana de contexto que hemos creado para que GPT-4 sepa como actuar cuando le presentemos un motivo de revisión fáctica de un recurso de suplicación.

La primera parte del 'prompt' de la ventana de contexto serían todos los elementos anteriormente señalados, a saber, el rol de juez, la tarea de resolver un recurso, la forma de una sentencia, qué son los hechos probados, etc.

La parte esencial aparece a continuación, y parte de las distintas técnicas anteriormente expuestas. Dado que la ventana de contexto utilizada ocupa unas 6 páginas, mostraremos alguna de sus partes para ejemplificar cómo se han utilizado las técnicas de prompting anteriormente indicadas.

Para llevar a cabo la ventana de contexto se le ha dicho a la IA que hay unas reglas que debe seguir, paso a paso ('Zero Shot Chain of Thought'), para resolver el recurso.

"Para resolver el recurso tienes que pensar paso a paso y tienes que tener en cuenta las siguientes reglas. Las reglas son las siguientes:"

Esas reglas son tomadas del libro 'El recurso de suplicación: la revisión de los hechos probados'¹⁰, simplificando cada una de ellas y pasando de lo más sencillo a lo más complicado. Así mismo, se han separado cada una de las reglas por '####', siguiendo la guía de mejores prácticas¹¹ publicada por la propia OpenAI, dado que estos símbolos permiten a la IA diferenciar las distintas ideas e instrucciones.

Como ejemplo tenemos los siguientes:

"####

En primer lugar, tienes que analizar si el Hecho Probado que se propone es un 'hecho probado indirecto'. Si es un hecho probado indirecto no se admite la revisión. Un ejemplo de 'hecho probado indirecto' es: "En el documento X pone 'Lorem ipsum dolor sit amet, consectetur adipiscing elit...'. "El informe Y dispone lo siguiente 'Ut enim ad minim veniam, quis nostrud exercitation ullamco...'"

Un 'hecho indirecto' consiste en la afirmación de que una prueba obrante en las actuaciones tiene un determinado contenido.

Esta revisión debe ser rechazada, porque la inclusión en los 'hechos probados' de hechos indirectos induce a confusión por falta de

¹⁰ Molins García-Atance, J. (2005). El recurso de suplicación: la revisión de los hechos probados. Thomson Reuters Aranzadi. ISBN 84-9767-858-3.

¹¹ Shieh, J. (2023, March). Best practices for prompt engineering with OpenAI API: How to give clear and effective instructions to GPT-3 and Codex. Disponible en: <https://help.openai.com/en/articles/6654000-best-practices-for-prompt-engineering-with-openai-api>

claridad, infringiendo así el mandato que contiene el art. 218.1 LEC, que es instrumental respecto al deber constitucional de motivación ex art. 24 y 120.3 CE, pues no se sabe a ciencia cierta si da por acreditado un hecho o el contenido de un medio de prueba, que es cosa bien distinta.

#####

Otro ejemplo de regla a seguir por la IA sería:

#####

En octavo lugar tienes que analizar si el Hecho Probado que se propone se basa en la misma prueba que ha tenido en cuenta el juzgador de la Sentencia recurrida para poner un Hecho Probado distinto. En este caso, no se admite la revisión. Un ejemplo: En la Sentencia recurrida el Hecho Probado Primero dice 'El sol es amarillo' y para ello el Juzgador de la Sentencia recurrida se basa en el documento nº 1. El recurrente pretende modificar el Hecho Probado Primero para que diga 'El sol es azul' y para ello se basa en el mismo documento nº 1 que ya había valorado e interpretado el juzgador de la Sentencia recurrida.

Esta revisión debe ser rechazada, porque la valoración de los documentos o periciales corresponde al 'juzgador de instancia' (el que ha dictado la sentencia recurrida) dado que su criterio imparcial prevalece sobre la interpretación parcial que pueda dar al mismo documento el recurrente (salvo un error manifiesto del juzgador al valorar la prueba, que has de mencionar al final del dictamen).

No cabe invocar el mismo documento o pericia en que se ha basado la sentencia recurrida para fijar el hecho y explicar que, aunque dicho medio de prueba no contradice el relato fáctico de instancia, el documento o pericia carece de eficacia probatoria.

El recurso debe articularse al revés. Es necesario invocar un documento o pericia distinto que demuestre el error probatorio de instancia. En definitiva, recurrente deberá identificar un documento o pericia que afirme algo distinto del hecho impugnado y demuestre así su equivocación.

#####

Como se puede ver, se trata de instrucciones o reglas muy sencillas, claras, en las que se dice lo que hay que hacer y lo que tiene que hacer constar cuando argumente su respuesta, utilizando ejemplos y la forma en que han de ser resueltos ('Chain of Thought').

Una vez creado todo el contexto, a saber, todas las reglas que la IA habrá de emplear para resolver la revisión fáctica planteada, habremos de indicar a la IA el final del contexto. A saber, reiteramos la forma en que debe de realizar la tarea y le presentamos la tarea:

####

Estas son todas las reglas que tienes que seguir. Piensa paso a paso, analiza el texto del recurso, las palabras usadas en el texto alternativo propuesto, y en los documentos que se proponen.

Este es el recurso que tienes que resolver:"

El final de la ventana de contexto es la presentación del recurso. Hay que tener en cuenta que la ventana de contexto se sitúa en el código de la siguiente manera:

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

La ventana de contexto se encuentra dentro del 'role' de 'system', es decir, le estamos diciendo al sistema, a GPT-4, cuáles son las reglas, de tal manera que cuando el 'role' 'user' introduzca su recurso, el 'role' 'assistant' responderá con la estimación o desestimación de la revisión fáctica. Dicha respuesta es la que se mostrará en la ventana de la derecha. Un ejemplo ilustrativo sería el siguiente:

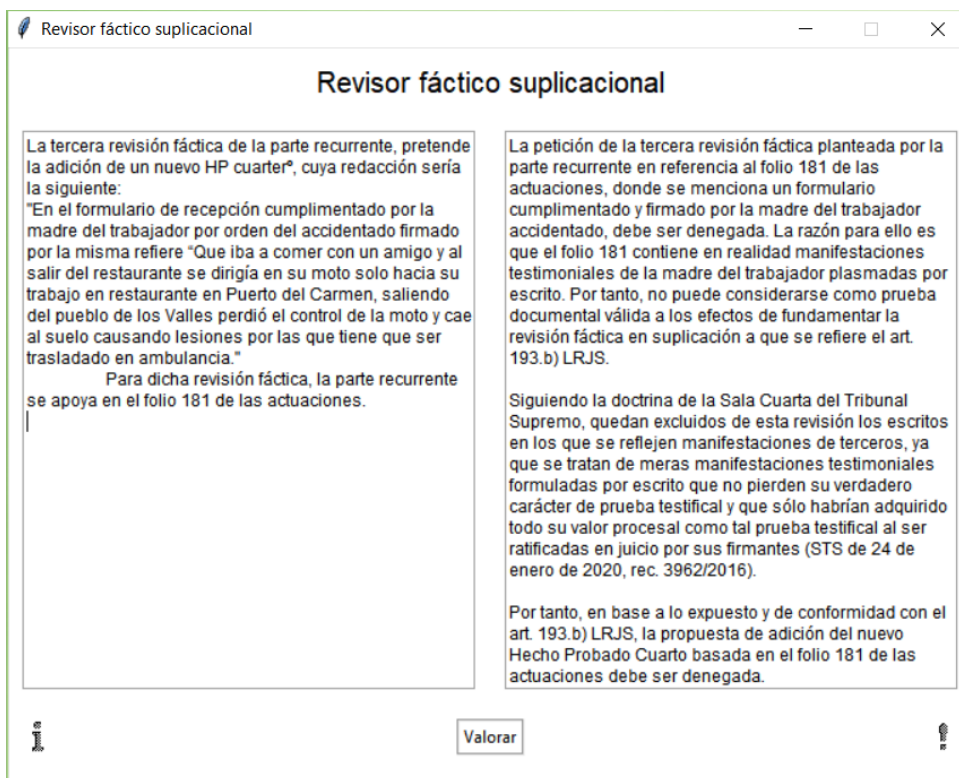


Imagen 4: La interfaz gráfica del programa en Tkinter muestra los cuadros de texto de entrada y salida, en la entrada se muestra una revisión fáctica, en la salida el dictamen de la IA denegando la revisión fáctica por ser una testifical documentada.

2.3. Proceso de corrección y aprendizaje

El programa desarrollado incorpora un mecanismo de corrección y aprendizaje que permite mejorar la precisión y eficiencia de las respuestas generadas por GPT-4. En caso de que GPT-4 resuelva de manera incorrecta o imprecisa un recurso de suplicación, el usuario tiene la opción de corregir manualmente la respuesta utilizando el botón con

símbolo '!', a la derecha de la ventana, tras el cuadro de salida.

Al hacer clic en este botón, se abre una ventana adicional en la que el usuario puede proporcionar la respuesta correcta y una explicación sobre el error cometido por GPT-4. Esta información es enviada al modelo al hacer clic en el botón "Enviar corrección". A partir de la corrección y la explicación proporcionadas, GPT-4 genera un resumen y crea una regla que se incluye en el contexto proporcionado al modelo. Esta regla permite que GPT-4 no cometa el mismo error en situaciones futuras similares, mejorando su precisión y adaptabilidad.

Es muy importante la forma en que se ha desarrollado esta forma de corrección de errores. Como ya se señaló anteriormente, una forma de 'prompting' es el 'Generated Knowledge' o enfoque de conocimiento generado, que consiste en generar conocimiento a partir de un modelo de lenguaje, y luego proporcionar dicho conocimiento como aporte adicional al responder una pregunta. Así mismo, también se señaló previamente que uno de los mayores problemas que tenía GPT-4 y las IA conversacionales de generación de texto era la memoria a largo plazo, es decir, cerrada la sesión, todo el contexto en el que se había desarrollado la conversación anterior se pierde. Los ingenieros de IAs podrían hacer ventanas de contexto mucho más grandes, y esto seguiría siendo así: cada vez que comenzaras una nueva conversación con GPT-4, estarías empezando desde cero. Como señala Jacob Stern en el artículo ya citado, "When booted up, it is borna new"¹².

Por lo tanto, difícilmente podría aprender de las sesiones y correcciones que pudieran hacer los usuarios. Una solución es la propuesta en el presente trabajo. La ventana de contexto que hemos dado al programa ha de ser lo suficientemente clara y concreta como para poder resolver las revisiones fácticas, pero al mismo tiempo, tiene que ser lo suficientemente escueta para dejar espacio a la 'retroalimentación' de los usuarios. Actualmente, la ventana de contexto en GPT-4 8k es de unas 12 páginas, habiendo usado 6 para redactar todas las reglas e instrucciones necesarias. Consecuentemente, tenemos aún 6 páginas para integrar las correcciones.

Sin embargo, las correcciones pueden ser muy extensas. Como se señaló antes, la 'Chain of Thought' o cadena de pensamiento expone los problemas y las soluciones, para que la IA pueda partir de ese conocimiento y responder de la forma correcta y a la manera deseada. Si la forma de retroalimentar a la IA, en caso de error en su dictamen, fuera poner toda la revisión y la explicación del usuario, tendríamos un texto como el siguiente:

«Revisión: La tercera revisión fáctica de la parte recurrente, pretende la adición de un nuevo HP quarter°, cuya redacción sería la siguiente:

¹² "Cuando se enciende, nace de nuevo". Stern, J. (2023) GPT-4 Has the Memory of a Goldfish. The Atlantic. Disponible en: <https://www.theatlantic.com/technology/archive/2023/03/gpt-4-has-memory-context-window/673426/>

"En el formulario de recepción cumplimentado por la madre del trabajador por orden del accidentado firmado por la misma, refiere "Que iba a comer con un amigo y al salir del restaurante se dirigía en su moto solo hacia su trabajo en restaurante en Puerto del Carmen, saliendo del pueblo de los Valles perdió el control de la moto y cae al suelo causando lesiones por las que tiene que ser trasladado en ambulancia."

Para dicha revisión fáctica, la parte recurrente se apoya en el folio 181 de las actuaciones.

Retroalimentación: *La respuesta es errónea, estamos ante una testifical documentada, dado que la declaración que se contiene se refiere a una persona, que podría haber despedido en el acto del juicio, limitando con ello la posible contradicción y la posibilidad de que la parte contraria formulara preguntas.»*

Este texto suponen unos 310 tokens, si contamos con la mitad de los 8.192 tokens disponibles para la ventana de contexto, una retroalimentación así limitaría mucho la capacidad de aprendizaje de la IA.

Por ello, partiendo de la técnica de 'Generated Knowledge', lo que se hace es tomar ambos textos y pasárselos a GPT-4, con un 'prompt' preciso para que, de esa propuesta de revisión y de esa corrección, saque una 'regla' que incluir en la ventana de contexto, al final del archivo.

La ventana de contexto se configura como un archivo de texto (.txt), lo que facilita su edición, dicho archivo contiene el 'prompt' y al final del mismo, tras todas las instrucciones, contempla lo siguiente:

#####

En último lugar, revisa las siguientes reglas, son precisiones de las reglas anteriores, y están basadas en fallos previos que has cometido:

##"

Al ser un archivo de texto, el propio programa, con cada iteración puede editar dicho archivo, de tal manera que efectuada una corrección, si inmediatamente se vuelve a plantear el mismo recurso, la solución que dará será distinta, a saber, atendiendo a la nueva regla creada.

El código es el siguiente:

```

def abrir_ventana_feedback():
    def enviar_feedback():
        feedback = cuadro_feedback.get("1.0", "end-1c")
        texto_original = left_text.get("1.0", "end-1c")
        regla = openai.ChatCompletion.create(
            model="gpt-4",
            messages=[
                {"role": "system", "content": "Eres una IA que asume
el rol de juez que resuelve recursos de suplicación. El siguiente
texto es una corrección ante un recurso que se te ha planteado y lo
has resuelto mal. La corrección corresponde al usuario, que ha
considerado que el resultado que tú has dado es incorrecto. Estudia
cuál es el recurso y la corrección del usuario, extrae de este texto
una regla clara, completa y precisa que puedas utilizar a partir de
ahora para resolver los recursos. Escríbela de forma que tú la
comprendas y puedas evitar cometer el mismo error la próxima vez.
Esta regla se incluirá dentro de un archivo donde están el resto de
las reglas que utilizarás para resolver los recursos. La estructura
es 'Regla:' y la explicación. Escribe la regla a partir del
siguiente texto:"},
                {"role": "user", "content": f"Petición de revisión:
{texto_original}\n Solución resumida: {feedback}"},
            ]
        )
        with open("revisor_factico.txt", "a", encoding='utf-8') as
f:
            f.write(regla['choices'][0]['message']['content']) +
"\n#\n")
        ventana_feedback.destroy()
        ventana_feedback = tk.Toplevel(root)
        ventana_feedback.title("Retroalimentación")
        ventana_feedback['bg'] = "#ffffff"
        cuadro_feedback = tk.Text(ventana_feedback, wrap=tk.WORD,
width=45, height=10, font=("SegoeUI", 9), relief='groove', bd=2)
        cuadro_feedback.pack(expand=True, fill=tk.BOTH, padx=10)
        botón_cerrar = tk.Button(ventana_feedback, text="Cerrar",
command=enviar_feedback, bg="otón", activebackground="light grey",
relief="otón")
        botón_cerrar.pack(pady=10, padx=10)

```

Como se puede apreciar en el código, el 'prompt' que se le da a la IA para que cree una nueva regla en atención a la corrección efectuada es el siguiente:

"Eres una IA que asume el rol de juez que resuelve recursos de suplicación. El siguiente texto es una corrección ante un recurso que se te ha planteado y lo has resuelto mal. La corrección corresponde al usuario, que ha considerado que el resultado que tú has dado es incorrecto. Estudia cuál es el recurso y la corrección del usuario, extrae de este texto una regla clara, completa y precisa que puedas utilizar a partir de ahora para resolver los recursos. Escríbela de forma que tú la comprendas y puedas evitar cometer el mismo error la próxima vez. Esta regla se incluirá dentro de un archivo donde están el resto de las reglas que utilizarás para resolver los recursos. La estructura es 'Regla:' y la explicación. Escribe la regla a partir del siguiente texto:"

El resultado, por lo tanto, será la creación de una regla que la IA entienda, que se añada a la ventana de contexto y que le sirva para no volver a caer en el mismo error; se salva con ello la situación de que "[w]hen booted up, it is borna new", dado que ahora, cada vez que se inicie, habrá guardado cada regla y cada corrección. A este 'prompt' habría que añadirle una última indicación, a saber:

"Revisa las reglas anteriores, si alguna es equivalente a la que vas a añadir, no la generes"

Este último prompt viene a evitar la duplicidad de reglas, de tal manera que las 6 páginas que quedan (o 4.096 tokens) sean utilizadas de manera eficiente para mejorar el sistema de resolución de revisiones fácticas.

Un ejemplo de regla creada por la propia IA sería el siguiente:

"Regla: Cuando se presente una revisión fáctica basada en declaraciones de personas contenidas en un documento, no admitirla como prueba válida, ya que se trata de una testifical documentada y no es admisible en el proceso."

La regla anterior, generada por la propia IA, supone únicamente 72 tokens, lo que supone la posibilidad de incluir en la ventana de contexto unas 50 nuevas reglas generadas a partir de las correcciones de los usuarios.

2.4. Integración de la API de OpenAI

Para utilizar GPT-4 en el programa en Tkinter, se requiere la integración de la API de OpenAI. Esta API permite la comunicación entre el programa y el modelo GPT-4, facilitando el envío de información y la recepción de respuestas generadas por el modelo.

La integración de la API de OpenAI en el programa se realiza mediante la biblioteca openai de Python, que ofrece funciones y métodos para interactuar con el modelo de manera sencilla y eficiente. Al utilizar la API de OpenAI, el programa puede enviar el contexto, la información del recurso de suplicación y las correcciones proporcionadas por el usuario a GPT-4. A su vez, GPT-4 genera respuestas y reglas que se incorporan en el contexto, mejorando su capacidad para resolver recursos de suplicación de manera adecuada y conforme a las directrices legales.

2.5. Pruebas y evaluación del programa

Durante el desarrollo del programa en Tkinter, se llevaron a cabo pruebas y evaluaciones para determinar su eficacia en la resolución de recursos de suplicación y su capacidad para aprender de las correcciones proporcionadas por el usuario. Estas pruebas incluyeron la introducción de casos reales y simulados, así como la evaluación de las respuestas

generadas por GPT-4 en términos de precisión, coherencia y fundamentación jurídica.

Los resultados de las pruebas indicaron que, en general, GPT-4 es capaz de generar respuestas adecuadas y coherentes con el contexto proporcionado, resolviendo correctamente la mayoría de los recursos de suplicación presentados. Además, se observó que el mecanismo de corrección y aprendizaje implementado en el programa contribuye significativamente a mejorar la precisión de las respuestas generadas por GPT-4 en situaciones futuras similares.

Sin embargo, también se identificaron áreas de mejora y situaciones en las que GPT-4 aún puede presentar dificultades para comprender completamente el contexto o aplicar correctamente las directrices legales. Estos hallazgos destacan la importancia de continuar mejorando el programa y el modelo GPT-4 para garantizar una mayor precisión y eficiencia en la resolución de recursos de suplicación en el ámbito jurídico.

Resulta interesante mostrar alguna de las respuestas previo examen de las revisiones propuestas:

1ª Revisión
<p>La segunda revisión fáctica de la parte recurrente, pretende la modificación del HP 8°. En la Sentencia de Instancia, el HP 9° tiene la siguiente redacción:</p> <p><i>"La cámara que captó las imágenes de la agresión de D. Ernesto a un ciudadano estaba situada en la fachada de un edificio y orientada a la puerta por la que los vehículos entran y salen del inmueble (Prueba documental número 11 a 14 de la parte demandante)"</i></p> <p>La redacción que se propone sería la siguiente:</p> <p><i>"La cámara que captó las imágenes de la agresión de D. Ernesto a un ciudadano estaba situada en la fachada de un edificio y orientada a la puerta por la que los vehículos entran y salen del inmueble. No consta un dispositivo informativo en lugar suficientemente visible identificando la cámara en cuestión (Prueba documental número 11 a 14 de la parte demandante)"</i></p> <p>Para dicha revisión fáctica, la parte recurrente se apoya en los documentos 11 a 14 de la recurrente (folios 176 a 180 de autos).</p>
Dictamen
<p>Al analizar la segunda revisión fáctica propuesta por la parte recurrente, es necesario tener en cuenta las reglas previamente enumeradas y aplicarlas al caso concreto.</p> <p>En este caso, la parte recurrente solicita la modificación del Hecho Probado 9°, proponiendo una redacción alternativa que añade una frase al final del original: "No consta un dispositivo informativo en lugar suficientemente visible identificando la cámara en cuestión".</p> <p>La parte recurrente fundamenta su petición de revisión en la prueba documental números 10 a 13, aportados por la parte demandante en la instancia (folios 146 a 149 de autos). Se cumplen así las reglas 12 y 14, en cuanto a la propuesta de texto alternativo y la cita pormenorizada del documento en el que se basa la revisión.</p> <p>No obstante, cabe hacer referencia aquí a la regla 4, según la cual no procede la inclusión de hechos probados negativos en la revisión fáctica. La frase propuesta por la recurrente tiene un carácter negativo al afirmar que "no consta" un elemento (un dispositivo informativo visible). Dicha afirmación no aporta un hecho concreto al relato histórico de la sentencia recurrida y no cumple con función procesal alguna.</p>

Por tanto, habida cuenta de las reglas anteriormente enumeradas y aplicándolas a la revisión fáctica propuesta, procede denegar la petición de modificación del Hecho Probado 8° por ser un hecho probado negativo que no cumple función procesal alguna y no aporta ningún elemento adicional al relato histórico de la sentencia recurrida.

2ª Revisión

La octava revisión fáctica de la parte recurrente, pretende la modificación del HP 8°. En la Sentencia de Instancia, el HP 8° tiene la siguiente redacción:

"La cámara captó como D. Pedro agredía a un ciudadano sin una provocación previa. (Prueba documental número 10 a 13 de la parte demandante)"

La redacción que se propone sería la siguiente:

" La cámara captó como D. Pedro se defendía de la agresión de un ciudadano. (Prueba documental número 10 a 13 de la parte demandante)"

Para dicha revisión fáctica, la parte recurrente se apoya en los documentos 10 a 13 de la recurrente valorados por el juez a quo.

Dictamen

Ante la revisión fáctica propuesta en el recurso referente al Hecho Probado 8°, debemos analizarla de acuerdo con las reglas mencionadas previamente:

1.- La recurrente ha propuesto una redacción alternativa específica para el Hecho Probado 8°. Por lo tanto, cumple con esta regla.

2.- Ha identificado tanto el hecho afirmado en la sentencia de instancia (D. Pedro agrediendo a un ciudadano sin provocación previa) como el hecho que considera contrario a lo acreditado (D. Pedro defendiéndose de la agresión de un ciudadano).

3.- La recurrente ha citado de manera pormenorizada los documentos en los que basa su pretensión revisora de los Hechos Probados: la prueba documental números 10 a 13 de la parte demandante.

Sin embargo, cabe destacar lo siguiente:

4.- La recurrente pretende modificar el Hecho Probado 8° basándose en las mismas pruebas documentales (números 10 a 13 de la parte demandante) que fueron valoradas por el juez a quo en la sentencia de instancia.

Siguiendo las reglas mencionadas anteriormente, el Tribunal Constitucional ha declarado que el derecho a la tutela judicial efectiva no ampara la valoración de la prueba conforme a las pretensiones ejercitadas por la recurrente. La parte recurrente no ha presentado una prueba diferente a las ya valoradas por el juez de instancia que demuestre su posición. Por lo tanto, no procede en este caso la modificación del Hecho Probado 8° basándose en las mismas pruebas documentales ya valoradas por el juez de instancia.

Por todo lo anterior, la petición de modificación del Hecho Probado 8° debe ser denegada, ya que la recurrente pretende basarse en las mismas pruebas documentales valoradas por el juez a quo, sin presentar una prueba diferente que demuestre el error en los Hechos Probados de instancia.

3ª Revisión

La segunda revisión fáctica de la parte recurrente, pretende la adición de un nuevo HP 24°, cuya redacción sería la siguiente:

"Desde el 10 de agosto de 2021 hasta el 20 de diciembre de 2021 el demandante percibió las siguientes cantidades salariales:

21 días de Agosto de 2021: 1.593,42 EUROS.

Septiembre de 2021: 1.807,83 EUROS.

Octubre de 2021: 1.869,18 EUROS.

Noviembre de 2021: 1.865,96 EUROS.

Diciembre de 2021: 1.510,78 EUROS.

12,60 días de vacaciones devengadas y no disfrutadas: 733,98 EUROS"

Para dicha revisión fáctica, la parte recurrente se apoya en los

documentos 330 a 336.

Dictamen

Para resolver la segunda revisión fáctica de la parte recurrente que solicita la adición de un nuevo Hecho Probado (HP) 24° con la redacción propuesta, hemos de analizar si cumple con las reglas mencionadas previamente.

1. Se ha ofrecido un texto alternativo concreto que refleja de manera pormenorizada las cantidades salariales percibidas por el demandante durante el período señalado. En este caso, la revisión cumple con la regla 12.

2. Además, la recurrente ha indicado de manera precisa y clara cuál es el hecho afirmado u omitido, que en este caso es una omisión: las cantidades salariales percibidas por el demandante. Por lo tanto, la revisión cumple con la regla 13.

3. La parte recurrente ha citado de manera pormenorizada los documentos, del 330 al 336, en los cuales se basa para fundamentar la solicitud de revisión fáctica. Por ende, la revisión cumple con la regla 14.

Lo que debemos analizar a continuación es si los documentos citados cumplen con las reglas y requisitos para la revisión de los Hechos Probados. Es decir, únicamente pueden utilizarse como prueba documental o pericial, y deben demostrar el error probatorio del juez.

Dado que no se han mencionado detalles específicos sobre el contenido de los documentos del 330 al 336, no es posible resolver el recurso sin analizar y evaluar si estos documentos cumplen con todos los requisitos anteriormente expuestos.

Por lo tanto, en este caso, sería necesario examinar los documentos proporcionados por la parte recurrente para determinar si se ajustan a las reglas y requisitos correspondientes para la revisión de los Hechos Probados.

En consecuencia, a falta de información adicional sobre el contenido de los documentos citados, la resolución del recurso dependerá de si los documentos aportados cumplen con los requisitos necesarios para fundamentar la revisión fáctica en el recurso de suplicación.

4ª Revisión

La única revisión fáctica de la parte recurrente, pretende la adición de un nuevo HP 12°, cuya redacción sería la siguiente:

"Son funciones del auxiliar educativo, nueva denominación de la categoría de cuidador/a, Grupo IV, las siguientes:

- Controlar y atender la higiene y aseo personal del alumno o residente, durante a la Dirección aquellas anomalías observadas en la higiene y aseo de su estancia en el Centro/Residencia.

- Notificar los alumnos o residentes tras la recepción de los mismos. Colaborar en la aplicación del programa de control de esfínteres

- Dar de comer o suministrar la alimentación a los alumnos o residentes que no puedan hacerlo por si mismos, salvo en casos que requieran cuidados especiales de tipo médico-sanitario.

Tareas éstas de índole técnico-práctico y de cuidado de prestación asistencial, en la que un docente es obvio que no participa."

Para dicha revisión fáctica, la parte recurrente se apoya en el art. 26 de la Orden de la Consejería de Educación del Gobierno de Canarias de 13.12.10, BOC de 22.12.10, obrante al folio 730 y siguientes de las actuaciones, en el que se definen las funciones de las categorías profesionales.

Dictamen

Tras examinar el recurso de suplicación planteado, para resolverlo debemos analizar si se ajusta a las reglas establecidas en el marco legal y jurisprudencial mencionado previamente.

En este caso, se pretende la adición de un nuevo Hecho Probado (HP) 12° en la sentencia recurrida que describa las funciones del auxiliar educativo o cuidador/a, y para ello se invoca como prueba documental el art. 26 de la Orden de la Consejería de Educación del Gobierno de Canarias de 13.12.10, BOC de 22.12.10, obrante al folio

730 y siguientes de las actuaciones.

El recurso cumple en principio con las reglas relativas a la propuesta de un texto concreto que se pretende incorporar en los Hechos Probados y la identificación específica del documento en que se basa la parte recurrente. La descripción de las funciones de la categoría profesional estaría relacionada con un hecho relevante para el caso presente.

Sin embargo, es importante analizar si el hecho que se pretende incorporar contiene aspectos de valoración jurídica o calificación jurídica que deberían ser abordados en los Fundamentos de Derecho y no en los Hechos Probados. En este sentido, el propuesto HP 12º parece describir un conjunto de funciones que en su mayoría podrían considerarse hechos, pero también incluye algunas afirmaciones que podrían implicar una calificación jurídica. Es el caso de la frase *"Tareas éstas de índole técnico-práctico y de cuidado de prestación asistencial, en la que un docente es obvio que no participa."* Esta afirmación incluye una conclusión jurídica sobre el tipo de tareas y la exclusión de los docentes en la realización de esas funciones, lo que podría ser tratado en los Fundamentos de Derecho.

Por lo tanto, teniendo en cuenta lo anterior, se podría admitir la revisión fáctica solicitada, pero con la modificación de eliminar la parte que contiene la valoración jurídica, y en consecuencia, incorporar en los Hechos Probados el nuevo HP 12º, redactado de la siguiente manera:

"Son funciones del auxiliar educativo, nueva denominación de la categoría de cuidador/a, Grupo IV, las siguientes:

- Controlar y atender la higiene y aseo personal del alumno o residente, durante a la Dirección aquellas anomalías observadas en la higiene y aseo de su estancia en el Centro/Residencia.

- Notificar los alumnos o residentes tras la recepción de los mismos. Colaborar en la aplicación del programa de control de esfinteres

- Dar de comer o suministrar la alimentación a los alumnos o residentes que no puedan hacerlo por si mismos, salvo en casos que requieran cuidados especiales de tipo médico-sanitario."

De esta forma, se admitiría parcialmente el recurso de suplicación presentado y se procedería a la modificación del relato fáctico en la sentencia recurrida conforme a la propuesta presentada, a excepción de la eliminación de la parte que contiene valoración jurídica, que deberá ser tratada en los Fundamentos de Derecho si corresponde.

5ª Revisión

La segunda revisión fáctica de la parte recurrente, pretende la adición de un nuevo HP 5º, cuya redacción sería la siguiente:

"Que en el Ayuntamiento de Tías presta servicios Dña. Mercedes Rodríguez Rodríguez, realizando las mismas funciones que la actora recogidas en el hecho probado segundo según se recoge en el fundamento de derecho cuarto de la sentencia y se le retribuye como una auxiliar administrativa (auxiliar de biblioteca) encuadrada en el Grupo C2 Nivel 18"

Para dicha revisión fáctica, la parte recurrente se apoya en el documento nº 11 (folios 160 y 161 de las actuaciones), en el documento nº 10 (folios 139 y 140) y en la testifical de la propia trabajadora Dña. Mercedes Rodríguez Rodríguez.

Dictamen

Para resolver la segunda revisión fáctica planteada por la parte recurrente en relación con la adición de un nuevo Hecho Probado 5º, debemos analizar si cumple con las reglas establecidas para la revisión de hechos probados en un recurso de suplicación.

La parte recurrente propone un texto concreto para figurar como Hecho Probado 5º en la sentencia, cumpliendo con la regla 12. Asimismo, cumple con regla 13, al indicar de manera precisa y clara cuál es el

hecho que considera omitido.

Sin embargo, la parte recurrente apoya su revisión fáctica en la declaración testifical de Dña. Mercedes Rodríguez Rodríguez, lo cual incumple la regla 9, ya que para la revisión de hechos probados solamente pueden invocarse documentos o pericias.

En cuanto a los documentos citados, la parte recurrente hace referencia al documento nº 11 (folios 160 y 161) y al documento nº 10 (folios 139 y 140). Cumple con la regla 14, al citar de manera pormenorizada los documentos en los que basa su pretensión.

Para determinar si la pretensión es procedente, debemos evaluar si dichos documentos aportan información suficiente y distinta de la que ha sido apreciada por el juez de instancia en la sentencia recurrida, conforme a la regla 1.

Si los documentos nº 11 y nº 10 aportan pruebas claras y suficientes que respalden la afirmación propuesta, la revisión fáctica podría ser estimada y procedería la adición del nuevo Hecho Probado 5º en la sentencia recurrida.

De lo contrario, si los documentos no aportan pruebas concluyentes que apoyen la afirmación propuesta, la revisión fáctica no prosperaría, y la sentencia recurrida permanecería inalterada en cuanto a ese aspecto.

En conclusión, para resolver esta revisión fáctica, es necesario examinar el contenido de los documentos nº 11 y nº 10 y determinar si aportan pruebas suficientes y distintas de la que ha sido apreciada por el juez de instancia en relación con el nuevo Hecho Probado 5º propuesto. Si lo hacen, se procederá a la adición del mismo; en caso contrario, la revisión fáctica no prosperará. Sin embargo, en cualquier caso, debe desestimarse la referencia a la testifical, ya que no es admisible para la revisión fáctica en un recurso de suplicación.

Por todo lo expuesto, este cuarto motivo de revisión fáctica propuesto por la parte recurrente debe ser rechazado, al no cumplir con los requisitos necesarios ni fundamentarse en pruebas documentales o periciales que respalden su alegación.

Como se puede apreciar, el nivel de razonamiento y argumentación es correcto. El texto mantiene su coherencia desde el principio hasta el final. La IA es capaz de reconocer los fallos en las revisiones propuestas e incluso proponer redacciones alternativas que eludan 'hechos predeterminantes del fallo'. Por supuesto, sigue habiendo supuestos en los que falla, pero los mismos pueden ir siendo superados gracias a la capacidad de generación de nuevas reglas a partir de las correcciones de los usuarios.

2.6. Limitaciones

Las instrucciones dadas en la ventana de contexto a GPT-4 son básicamente todas aquellas reglas sobre la revisión fáctica que no precisarían ni del análisis de la prueba propuesta, ni del análisis del procedimiento en general, a saber, los límites de esta herramienta estarían en la literosuficiencia de la prueba propuesta y en la trascendencia la revisión propuesta al objeto de alterar el sentido del fallo.

Esta herramienta permite identificar:

- Los hechos probados indirectos
- Los hechos probados negativos
- Los hechos probados predeterminantes del fallo
- Los hechos probados notorios y conformes

- La revisión de hechos basada en la misma documental valorada por el juzgador/a de instancia
- La revisión de hechos basada en una testifical, interrogatorio, prueba de audio o videográfica
- La revisión de hechos que resultan de la apreciación por el juzgador/a de instancia de la testifical o interrogatorio practicado en juicio
- La revisión sin redacción alternativa
- La revisión sin especificar documento o pericial concreta
- La revisión basada en una nueva valoración global de la prueba

Los límites son importantes, dado que gran parte de las revisiones fácticas cumplen con los requisitos básicos, reseñados entre otras en la Sentencia 90/2022 del Tribunal Supremo, Sala Cuarta, de 1 de febrero (rec. 2429/2019); sin embargo, carecen de trascendencia para alterar el sentido del fallo. Este límite se encuentra en la propia configuración de los LLM basados en la tecnología transformer, y es que la ventana de contexto es muy difícil de ampliar.

Como hemos señalado reiteradamente a lo largo del presente artículo, GPT-4 8k ha ampliado la ventana de contexto a 8.192 tokens, unas 12 páginas, si bien GPT-4 32k, aún no publicada, lo ampliará a 32.768 tokens, unas 50 páginas. Esa es la memoria más impresionante que ha logrado un modelo transformer, el tipo de red neural en el que se basan todos los LLM más importantes. Un modelo como GPT-4 32k, sí podría leer el 90% de las sentencias de Juzgados de lo Social, dado que limitando el conocimiento a los Hechos Probados, Fundamentos de Derecho y Fallo (prescindiendo por ende de encabezados, antecedentes de hecho, pie de recurso) pocas sentencias se extienden más allá de 50 páginas. Un modelo como GPT-4 32k, podría analizar si la revisión propuesta, examinada la fundamentación jurídica y el fallo, tendría trascendencia para alterar el mismo, sin embargo, esa labor excede de una tarea de razonamiento cognitivo simple, y requiere una cierta valoración jurídica, lo cual, actualmente, está vedado a las IAs.

Los grandes modelos de lenguaje (LLM), como GPT-4 de OpenAI, son sistemas de inteligencia artificial basados en redes neuronales que han sido entrenados para procesar y generar texto. Aunque estos modelos pueden realizar tareas que involucran habilidades cognitivas como el razonamiento, la comprensión del lenguaje natural y la generación de texto, no pueden realizar tareas de cognición profunda en el sentido tradicional.

La cognición profunda generalmente se refiere a procesos cognitivos más complejos y de alto nivel, como la planificación, el pensamiento abstracto, la toma de decisiones, la solución de problemas y el aprendizaje. Los LLM pueden imitar ciertos aspectos de estos procesos a través de la generación de texto, pero su comprensión se basa en patrones y correlaciones aprendidos a partir de los datos de entrenamiento, en lugar de un conocimiento profundo y contextual de la realidad.

La capacidad de los LLM para "razonar" o "entender" se basa en patrones estadísticos en lugar de una comprensión consciente y conceptual del mundo. En resumen, aunque los LLM pueden realizar tareas que involucran habilidades cognitivas básicas, no se puede considerar que realicen cognición profunda en el sentido más completo y humano de

la palabra.

La valoración jurídica de un asunto o la interpretación de una norma jurídica son tareas que requieren habilidades cognitivas avanzadas y pueden considerarse tareas de cognición profunda en cierto sentido. Estas tareas implican el análisis y la comprensión de leyes y regulaciones, la evaluación de hechos y circunstancias, el razonamiento lógico y la toma de decisiones informadas.

Aunque los grandes modelos de lenguaje (LLM), como GPT-4, pueden ayudar en estas tareas proporcionando información relevante, generando posibles argumentos o sugiriendo interpretaciones basadas en patrones en el texto legal, su capacidad para realizar una cognición profunda en el ámbito jurídico es limitada. Los LLM no poseen la conciencia, el conocimiento contextual y la comprensión profunda que un juzgador/a humana tendría al enfrentarse a estos problemas.

2.7. Consideraciones éticas y de privacidad

Dado que el programa, aún tratándose de un concepto, estaría desarrollado para ser utilizado en el ámbito jurídico, es crucial abordar las consideraciones éticas y de privacidad relacionadas con el uso de la inteligencia artificial y el procesamiento del lenguaje natural en la resolución de recursos de suplicación. Estas consideraciones incluyen la protección de datos personales y la confidencialidad de la información proporcionada por el usuario.

Para garantizar la privacidad y seguridad de la información, el programa debería implementar medidas de protección de datos, como el cifrado y el almacenamiento seguro de la información. Además, es fundamental informar a los usuarios sobre cómo se utiliza su información y obtener su consentimiento antes de procesarla a través del programa y GPT-4. Igualmente, una utilidad como la anonimización del Kendoj resultaría útil para procesar los nombres propios antes de que los servidores de GPT-4 reciban esa información.

En cuanto a las consideraciones éticas, es importante reconocer que, aunque GPT-4 es una herramienta poderosa y precisa, no es infalible y puede cometer errores o generar respuestas inadecuadas en ciertas situaciones. Por lo tanto, los usuarios deberían utilizar el programa como una herramienta de apoyo y no como una solución definitiva, siempre ejerciendo su criterio profesional y responsabilidad en la toma de decisiones. Es decir, como un dictaminador de posibilidades, verificables por el usuario, más que como un sistema de decisión final.

En resumen, el desarrollo de un programa como el expuesto, utilizando GPT-4, representa un avance significativo en la aplicación de la inteligencia artificial y el procesamiento del lenguaje natural en el ámbito jurídico. La interfaz gráfica de usuario, la implementación de GPT-4 y el context-length ampliado, el proceso de corrección y aprendizaje, la integración de la API de OpenAI y las pruebas y evaluación del programa demuestran el potencial de esta herramienta para mejorar la eficiencia y la calidad en la toma de algunas soluciones

jurídicas.

3. Repercusiones

La aplicación de un programa como el propuesto, que utiliza GPT-4 para resolver la revisión fáctica en los recursos de suplicación podría generar diversas repercusiones en el ámbito jurídico. Estas implicaciones afectarían principalmente a la eficiencia y calidad de las resoluciones judiciales, al rol de los profesionales del derecho y la percepción de las ventajas y desventajas de la implementación de la inteligencia artificial en la revisión fáctica.

3.1. Impacto en la eficiencia y calidad de las resoluciones judiciales

Algunos de los efectos más notables podrían ser los siguientes:

Reducción de tiempos de resolución: La capacidad de GPT-4 para analizar rápidamente los recursos de suplicación y generar respuestas basadas en directrices legales y jurisprudenciales podría permitir una reducción considerable del tiempo de toma de decisiones en Tribunal Superiores. Es cierto que la revisión fáctica no es la más compleja de las tareas que un Magistrado de TSJ realiza al resolver un recurso de suplicación, pero una tarea cuya complejidad disminuye o se elimina siempre supone un ahorro de tiempo.

Mayor consistencia en las resoluciones: Al utilizar un marco jurídico y doctrinal consistente, GPT-4 garantiza que sus respuestas se basen en criterios homogéneos, lo que reduce la variabilidad en las resoluciones y disminuye la probabilidad de discrepancias entre decisiones similares.

Reducción de errores humanos: La inteligencia artificial de GPT-4 minimiza el riesgo de errores humanos en la interpretación y aplicación de las normas legales y la jurisprudencia. Además, los mecanismos de corrección y aprendizaje permiten al programa mejorar continuamente su precisión y adaptabilidad.

Acceso a información actualizada: Si se diera a GPT-4 acceso a una amplia base de datos de jurisprudencia y doctrina, como el CENDOJ, ya mediante una alimentación con la base de datos, o ya a través de la posibilidad de acceder a la misma (como hace BingChat), ello permitiría generar respuestas basadas en información actualizada y relevante. Esto mejoraría la calidad de las resoluciones y garantizaría que se consideraran los últimos cambios en el ámbito jurídico.

3.2. Cambios en el rol de los profesionales del derecho

La implementación de GPT-4 en la resolución de recursos y en otras posibles materias de resolución jurídica podría generar cambios en el rol de los profesionales del derecho. Estos cambios incluirían:

1. Mayor enfoque en la supervisión y corrección: Con GPT-4 asumiendo parte de la carga de trabajo en la revisión y resolución de cuestiones concretas del recurso de suplicación, los Magistrados podrían centrarse en la supervisión de las respuestas generadas por el programa y en la corrección de errores, garantizando que las decisiones tomadas sean coherentes con las normas legales y jurisprudencia aplicables. Esto supondría un ahorro de tiempo para los Magistrados y su empleo en otras partes del recurso más complejas (art. 193.a) y c) de la LRJS), y ello en virtud del problema P frente a NP¹³, conforme al cual es más difícil encontrar una solución a un problema que comprobar si esa solución es correcta. Los problemas P (polinómicos) son los que se pueden resolver en un tiempo razonable. Los problemas NP (no deterministas en tiempo polinómico) son aquellos que, aunque sea difícil encontrarles solución, una vez hallada se puede comprobar en un tiempo razonable su corrección. Consecuentemente, verificar que la IA ha resuelto correctamente la estimación o desestimación de una revisión fáctica conllevaría un tiempo menor que resolverlo y motivarlo directamente.
2. Desarrollo de habilidades en tecnología jurídica: La implementación de GPT-4 en el ámbito jurídico requeriría que los profesionales del derecho adquirieran habilidades en el uso de tecnologías emergentes y en la integración de estas herramientas en sus prácticas diarias. Esto implicaría un cambio en el perfil de habilidades requeridas para los abogados, jueces y otros profesionales del derecho.
3. Cambio en la toma de decisiones: La utilización de GPT-4 en la resolución de recursos de suplicación podría modificar la forma en que los Jueces o Magistrados tomaran sus decisiones. Al depender en gran medida de la inteligencia artificial para generar respuestas, los Jueces y Magistrados podrían verse influidos por las soluciones propuestas por GPT-4, lo que podría afectar su criterio y la forma de abordar los litigios.

3.3. Desventajas de la implementación de la IA en la revisión fáctica

La implementación de GPT-4 en la revisión fáctica podría generar una dependencia excesiva de la tecnología, lo que podría afectar al criterio y la toma de decisiones de los profesionales del derecho.

¹³ Ercilla García, J. (2020). Tribunales virtuales y procedimiento online: Solución de contingencia ante pandemias o evolución necesaria. Revista Aranzadi Doctrinal, (5), 1889-4380.

Así mismo, los modelos de inteligencia artificial, como GPT-4, pueden estar sujetos a sesgos presentes en los datos de entrenamiento, lo que afectaría a la calidad y equidad de las respuestas generadas.

Por último, a pesar de su avanzada capacidad de procesamiento del lenguaje natural, GPT-4 puede enfrentar dificultades en la interpretación de normas legales y jurisprudencia complejas o ambiguas, lo que puede generar errores o inconsistencias en sus respuestas.

4. Posibilidades futuras y evolución del programa

La aplicación de un programa en Tkinter que utilice GPT-4 para resolver la revisión fáctica en los recursos de suplicación es solo el comienzo del potencial que ofrece la IA en el ámbito jurídico. A medida que la tecnología avanza y se integra más en el ámbito legal, se pueden explorar diversas posibilidades futuras y mejoras en el programa y en su aplicación a otras áreas del derecho y de la resolución jurídica, sólo habría que buscar qué elementos son susceptibles de ser resueltos a través de tareas cognitivas básicas. En esta sección, analizamos algunas de las posibles evoluciones del programa en cuanto a la precisión y capacidad de aprendizaje de GPT-4, su expansión a otras áreas del derecho y ámbitos jurisdiccionales, y la integración de tecnologías complementarias y avances en el procesamiento del lenguaje natural.

4.1. Mejoras en la precisión y capacidad de aprendizaje de GPT-4

El potencial de GPT-4 para mejorar la precisión y la capacidad de aprendizaje es considerable, lo que puede dar lugar a un programa aún más eficiente y efectivo en la resolución de recursos de suplicación. Algunas de las mejoras posibles incluyen:

Entrenamiento con conjuntos de datos más amplios y diversos: A medida que se disponga de más información jurídica, jurisprudencia y doctrina, GPT-4 podría beneficiarse del entrenamiento con conjuntos de datos más amplios y diversos, lo que aumentaría su capacidad para generar respuestas precisas y contextualizadas.

Algoritmos de aprendizaje profundo más avanzados: El desarrollo de algoritmos de aprendizaje profundo más avanzados y eficientes podría mejorar la capacidad de GPT-4 para aprender de sus errores y adaptarse a nuevas situaciones y cambios en el ámbito jurídico. GPT-3.5 fue lanzado el 30 de Noviembre de 2022, GPT-4 fue lanzado el 14 de Marzo de 2023. Según OpenAI, GPT-3.5 obtuvo en el SAT¹⁴ una puntuación correspondiente al percentil 82°, GPT-4, cuatro meses después, obtuvo en el SAT una puntuación equivalente al percentil 94°; lo mismo cabría decir del LSAT¹⁵, para el que el GPT-3.5 obtuvo un percentil 40°, mientras que GPT-4 obtuvo un percentil 88°; y por último y más significativo, en el

¹⁴ Prueba estandarizada ampliamente utilizada para la admisión a la universidad en los Estados Unidos. Con prueba de lectura, escritura, lenguaje y matemáticas.

¹⁵ Prueba estandarizada administrada por el Consejo de Admisión de la Facultad de Derecho (LSAC) para posibles candidatos a la escuela de derecho. Está diseñado para evaluar la comprensión lectora, así como la competencia en el razonamiento lógico y verbal.

Uniform Bar Exam¹⁶, GPT-3.5 obtuvo el percentil 10°, mientras que GPT-3 obtuvo el percentil 90°, es decir, en el examen de derecho, paso de estar entre los 10 peores, a estar entre los 10 mejores. El avance en los sistemas LLM en poco tiempo (GPT-3 es de mayo de 2020) se produce a pasos agigantados, por lo que la mejora en los algoritmos puede llevarnos, de no considerar posible tareas cognitivas más allá de las básicas, a atisbar probable tareas cognitivas profundas. Y con ello un cambio en la forma de administrar justicia.

Incorporación de mecanismos de retroalimentación y validación: La implementación de mecanismos de retroalimentación y validación que permitan a los profesionales del derecho evaluar y corregir las respuestas de GPT-4 podría mejorar su precisión y capacidad de aprendizaje.

4.2. Expansión

El éxito de los LLM ha supuesto que la integración de tecnología de IA en trabajos que hasta hace poco parecían insustituibles, tenga más visos de realidad. En un reciente artículo¹⁷ publicado por OpenAI, se analiza el potencial que GPT-4 podría tener para los empleos. Para efectuar dicho análisis parte de lo que denominan 'exposición', definiendo la misma como la evaluación de si, al usar un GPT (3.5, 4 o futurible superior) o un sistema basado en GPT, se lograría reducir al menos en un 50% el tiempo necesario para que una persona realice una tarea específica o complete una Actividad Laboral Detallada (DWA, por sus siglas en inglés).

En dicho estudio se dividen los trabajos en trabajos no expuestos, trabajos que podrían lograr reducir al menos un 50% del tiempo en realizar tareas usando GPT, y trabajos que lograrían dicha reducción pero adaptando GPT a sus necesidades concretas.

El resultado que ofrece el estudio podría resultar en un principio desalentador, dado que las profesiones más expuestas serían las correspondientes a trabajos relacionados con la programación y la redacción. Siendo así que la redacción es uno de los elementos esenciales de la labor jurisdiccional. Sin embargo, la redacción no es sino el reflejo material de algo inmaterial, que es lo que el propio artículo señala como trabajos no expuestos, a saber, el 'critical thinking', o pensamiento crítico, que es la esencia de toda labor jurisdiccional, dado que juzgar no es sólo aplicar el caso concreto al supuesto de hecho previsto en la norma, sino que implica una comprensión del mundo, una interpretación de la norma, de las voluntades, las intenciones etc... Un conjunto de elemento que constituyen las aptitudes que el artículo anteriormente señalado indica como labores menos expuestas, esto es, el pensamiento crítico. Ahora bien, el artículo es cauto, dado que señala

¹⁶ Examen previo a ser admitido en el colegio de abogados y obtener la licencia para ejercer la abogacía.

¹⁷ Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models. OpenAI, OpenResearch, University of Pennsylvania. Disponible en: <https://arxiv.org/pdf/2303.10130.pdf>

que “las ocupaciones que requieren estas habilidades son menos propensas a verse afectadas por los modelos de lenguaje actuales”, es decir, que no podemos sostener la misma opinión con los modelos de lenguaje que puedan llegar.

4.3. Integración de tecnologías complementarias y avances en el procesamiento del lenguaje natural con bases de datos jurídicas

A medida que la inteligencia artificial y otras tecnologías avanzan, las bases de datos jurídicas, como el CENDOJ, podrían beneficiarse de la integración de tecnologías complementarias y avances en el procesamiento del lenguaje natural.

Con la simple generación de una ventana de contexto de seis páginas, hemos conseguido un ‘humilde’ revisor fáctico suplicacional que podría aliviar cierta carga de trabajo a los Magistrados, verificando que el dictamen es correcto, la redacción adecuada y la fundamentación asumible. Los modelos como GPT-3 permiten ir más allá de las ventanas de contexto y llevar a cabo un ‘fine tuning’, esto es, la posibilidad de alimentar directamente el modelo con información. GPT-3 fue entrenada con 175.000 millones de parámetros de aprendizaje automatizado, sin embargo, no ha sido alimentado de manera concreta y específica con bases de datos jurídicas. Este ‘fine tuning’ o ajuste fino, permite crear modelos de GPT especializados en determinadas materias. Si con 6 páginas de reglas sencillas hemos sido capaces de crear un remedo de asistente judicial, las posibilidades que podría ofrecer hacer ‘fine tuning’ sobre GPT-4, alimentándolo con la base de datos del CENDOJ, serían inimaginables. Podríamos tener, no ya asistentes que resolvieran, sino asistente que sugirieran la redacción de resoluciones conforme a la jurisprudencia o doctrina constitucional más reciente.

Uno de los problemas que este ‘fine tuning’ planteaba era su coste. En un artículo publicado el 13 de Marzo de 2023, se indicaba que analistas y tecnólogos estimaban que el proceso crítico de entrenamiento de un gran modelo de lenguaje como GPT-3 de OpenAI podría costar más de cuatro millones de dólares¹⁸. Igualmente, señala que el modelo LLaMA (un LLM de Meta), tardó alrededor de 1 millón de horas de GPU en entrenar, lo que calculado sobre los costes que AWS cobra por este tipo de procesamientos, habría costado unos 2,4 millones de dólares. Un gasto así parece inasumible para la Administración de Justicia, sin embargo, el avance en este tipo de tecnologías nos ofrece alternativas como Alpaca 7B¹⁹, un modelo de lenguaje de seguimiento de instrucciones basado en el modelo LLaMA 7B de Meta, que permite crear modelos LLM locales, semejantes a GPT-3.5 (text-davinci-003), pero a un coste mucho menor, a saber, menso de 600 dólares. Así mismo, Alpaca 7B sería un proyecto abierto, destinado solo para investigación académica, estando su uso

¹⁸ Vanian, J., Leswing, K. (2023, March 13). ChatGPT and generative AI are booming, but the costs can be extraordinary. CNBC. Disponible en: <https://www.cnbc.com/2023/03/13/chatgpt-and-generative-ai-are-booming-but-at-a-very-expensive-price.html>

¹⁹ Taori, R., Gulrajani, I., Zhang, T., Dubois, Y., Li, X., Guestrin, C., Liang, P., & Hashimoto, T. B. (2023). Alpaca: A Strong, Replicable Instruction-Following Model. Stanford University. Disponible en: <https://crfm.stanford.edu/2023/03/13/alpaca.html>

comercial prohibido, lo que lo hace un candidato perfecto para la generación de LLMs de contenido jurídico basados en la enorme base de datos del CENDOJ.

5. Conclusiones

Una herramienta como la expuesta en las líneas anteriores no es sino una mera 'prueba de concepto', un ejercicio de hasta dónde se puede llegar con escasos recursos para atisbar lo que podría alcanzarse con una dedicación plena. La tecnología GPT-4 acaba de ser liberada y lo expuesto ut supra no es sino una de las posibilidades que ofrece. GPT-4 es multimodal, por lo que no sólo atiende al texto (los 'prompts') sino que también puede atender a las imágenes, lo que supondría la capacidad de ver imágenes, describir lo que habría en las mismas y expresarlo. Esta posibilidad, contemplada en GPT-4 pero aún no liberada, añadiría un soporte más a la posibilidad de la revisión fáctica, a saber, la de aquellas basadas en documentos no escritos susceptibles de percepción visual inmediata, las fotografías (documental ex art. 333 LEC).

En todo caso, la identificación en el ámbito de la administración de la justicia de todas aquellas tareas susceptibles de ser asistidas y automatizadas por herramientas de procesamiento de lenguaje natural vendría a aliviar en gran parte la carga de trabajo de Juzgados y Tribunales, al tiempo que permitiría centrar esfuerzos en labores de cognición más profunda que implicaran el uso del pensamiento crítico, habilidades estas que por ahora le están vedadas a las IAs.

6. Código fuente.

El código fuente de la aplicación está disponible en:

<https://github.com/xvi82/revisorfactico>

Su uso requiere tener una API Key de OpenAI.

7. Referencias

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models. OpenAI, OpenResearch, University of Pennsylvania. Disponible en: <https://arxiv.org/pdf/2303.10130.pdf>

Ercilla García, J. (2020). Tribunales virtuales y procedimiento online: Solución de contingencia ante pandemias o evolución necesaria. Revista Aranzadi Doctrinal, (5), 1889-4380.

Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., & Iwasawa, Y. (2022). Large Language Models are Zero-Shot Reasoners. Disponible en: <https://arxiv.org/abs/2205.11916>

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2022). Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. ACM Computing Surveys. Disponible en: <https://doi.org/10.1145/3560815>

Liu, J., Liu, A., Lu, X., Welleck, S., West, P., Bras, R. L., Choi, Y., & Hajishirzi, H. (2021). Generated Knowledge Prompting for Commonsense Reasoning. Disponible en: <https://arxiv.org/abs/2110.08387>

Molins García-Atance, J. (2005). El recurso de suplicación: la revisión de los hechos probados. Thomson Reuters Aranzadi. ISBN 84-9767-858-3.

Shieh, J. (2023, March). Best practices for prompt engineering with OpenAI API: How to give clear and effective instructions to GPT-3 and Codex. Disponible en: <https://help.openai.com/en/articles/6654000-best-practices-for-prompt-engineering-with-openai-api>

Stern, J. (2023) GPT-4 Has the Memory of a Goldfish. The Atlantic. Disponible en: <https://www.theatlantic.com/technology/archive/2023/03/gpt-4-has-memory-context-window/673426/>

Taori, R., Gulrajani, I., Zhang, T., Dubois, Y., Li, X., Guestrin, C., Liang, P., & Hashimoto, T. B. (2023). Alpaca: A Strong, Replicable Instruction-Following Model. Stanford University. Disponible en: <https://crfm.stanford.edu/2023/03/13/alpaca.html>

Vanian, J., Leswing, K. (2023, March 13). ChatGPT and generative AI are booming, but the costs can be extraordinary. CNBC. Disponible en: <https://www.cnbc.com/2023/03/13/chatgpt-and-generative-ai-are-booming-but-at-a-very-expensive-price.html>

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. arXiv preprint arXiv:1706.03762. Disponible en: <https://arxiv.org/abs/1706.03762>

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2022). Chain of Thought Prompting Elicits Reasoning in Large Language Models. Disponible en: <https://arxiv.org/abs/2201.11903>

