

Evaluación de la IA como herramienta durante el Trabajo Fin de Grado de ámbito fiscal, ¿complemento o sustituto del estudiante?

Gonzalo Díaz-Heredero López

diazheredero01@gmail.com

Universidad Complutense de Madrid. Departamento de Economía Aplicada, Pública y Política, Facultad de Ciencias Económicas y Empresariales. Campus Somosaguas, Pozuelo de Alarcón (Madrid).

Recibido: 20 de febrero de 2025

Aceptado: 16 de junio de 2025

Resumen

Este artículo expone la experiencia de un estudiante de Derecho-ADE en el desarrollo de un Trabajo de Fin de Grado (TFG) a través de inteligencia artificial (IA). La investigación, dentro del ámbito de fiscalidad internacional, se enfocó en identificar diferencias clave en Convenios de Doble Imposición (CDIs) entre España y siete países, comparándolos con el Modelo de Convenio Tributario de la Organización para la Cooperación y el Desarrollo Económicos (MC OCDE) de 2017. El artículo destaca cómo la IA complementa, pero no sustituye, el trabajo del estudiante, quien desempeñó un papel crucial en la validación y supervisión de los resultados generados por la herramienta. Se narra el proceso desde la perspectiva del alumno, de forma que sea útil para otros estudiantes que se planteen proyectos similares. También para que el profesorado universitario pueda complementar su visión ante documentos similares que reciban de su alumnado.

Palabras clave: inteligencia artificial, procesamiento del lenguaje natural, convenios de doble imposición, fiscalidad internacional, ingeniería del prompt, trabajo de fin de grado, docencia universitaria

Códigos JEL: H25, H26, C63.

1. INTRODUCCIÓN

La IA ha transformado diversos sectores, entre los que se encuentra la Educación, el Derecho y la Economía. La fiscalidad, como disciplina intermedia entre el Derecho y la Economía, ha sido igualmente impactada por estas innovaciones tecnológicas. Múltiples autores como Contreras (2024) o De Silva et al. (2024) han tratado la aplicación de estas avanzadas herramientas en el sector educativo, ya sea para la investigación o para la mejora de clases y docencia, mostrando los potenciales usos que esta puede tener.

Dentro del ámbito del estudio de los sistemas fiscales internacionales resulta ya obvio que la globalización y la digitalización han incrementado la complejidad de las relaciones económicas entre Estados, relaciones cuya tributación debe ser regulada a través de CDIs u otros tratados internacionales para evitar la elusión fiscal y la doble imposición. Estos fenómenos, cada vez

más cambiantes, requieren para la creación, modificación y análisis de dichos instrumentos, herramientas más avanzadas y eficientes frente a los métodos tradicionalmente usados.

Por ello, el Trabajo de Fin de Grado del que trata este artículo se centró en el uso de un modelo avanzado de Procesamiento Del Lenguaje Natural (PLN)¹ con el fin de identificar patrones no convencionales en convenios de doble imposición firmados por España con los socios comerciales cruciales, además de tratar de encontrar similitudes e implicaciones fiscales en sus disposiciones.

La razón detrás de la elección de estas herramientas para llevar a cabo el TFG no fue una decisión trivial, pues se basó en la experiencia previa del alumno en su uso académico desde el lanzamiento del modelo GPT-3.5 de OpenAI el 30 de noviembre de 2022. Desde entonces, en diferentes asignaturas relacionadas con las ciencias económicas -como Investigación Comercial, Análisis y Planificación Financiera, Análisis y Consolidación Contable, etc.- se entrenó al *chatbot* como tutor especializado en cada una de ellas, siendo capaz de simplificar, sintetizar o esquematizar el contenido de dichas asignaturas, además de ser capaz de resolver dudas y dar ejemplos para una mejor comprensión de ellas, permitiendo al alumno una mayor autonomía en su labor académica.

El objetivo de este artículo es mostrar cómo, a través de un conocimiento técnico y especializado previo de una determinada materia, se puede ampliar de forma considerable el ámbito de actuación del alumno a la hora de realizar un TFG, ahorrando recursos -tiempo, capital humano, costes económicos- tanto a él como a su tutor, aparte de permitirles centrarse en los aspectos en que ambos generan un verdadero valor añadido. Como consecuencia, se relegan a un segundo plano aspectos formales rutinarios -como el estilo de citas, ortografía o formato-, que, si bien son imprescindibles, no constituyen el núcleo conceptual del trabajo, siendo un complemento eficaz para ambas partes.

El artículo se detalla en tres secciones, en la primera se desarrollará la experiencia del alumno usando el *chatbot* en el TFG, siguiendo una exposición lineal y cronológica de los problemas y soluciones que fueron encontrando tanto tutor como alumno a lo largo de él, desde la búsqueda de información inicial hasta el entrenamiento y aplicación de la IA en el trabajo. Después se expondrán los resultados del trabajo, los objetivos conseguidos y las mejoras a llevar a cabo en futuras investigaciones para conseguir un mejor *output* del asistente. Por último, se darán una serie de conclusiones enumeradas y se discutirá la posible proyección futura de la herramienta en similares contextos académicos y docentes.

2. EXPERIENCIA DEL USO DE LA IA EN EL TFG

2.1. PRIMER ACERCAMIENTO E INVESTIGACIÓN

Una vez asignado el tema del TFG al alumno, al ser estudiante del doble grado Derecho-Business Management, tuvo que realizar dicho trabajo en inglés, acordando con el tutor el estudio de los Convenios de Doble Imposición de España al tratarse de una materia de ámbito fiscal.

También acordamos que queríamos hacer uso de los modelos de IA recientemente desarrollados, ya que el año anterior, el tutor fue profesor de las asignaturas de Tax System I y Tax System II, en la que nos encomendó una entrega de un ejercicio creado a través de IA con

el fin de hacerlo lo más parecido y robusto posible respecto a los ejercicios que él mismo ha usado para formarnos, teniendo que contrastar luego nosotros en clase las alucinaciones, errores y nivel de complejidad de dicho ejercicio “artificial”.

Con todo el conocimiento previo que el estudiante tenía acerca de estas herramientas se consiguió hacer un ejercicio lo suficientemente robusto a través de diferentes técnicas *-few shot learning, prompt engineering y context augmentation-*, siendo el ejercicio que menos errores y mayor similitud tuvo de la clase, realizado a través del modelo gratuito GPT-4, el cual tiene una distancia considerable de desarrollo respecto al último GPT-4o como veremos más adelante.

Aun poseyendo el alumno conocimientos básicos de este tipo de herramientas, la primera cuestión era buscar información profunda sobre los conceptos, pues no tenía formación específica sobre su funcionamiento ni sobre los principios técnicos que los rodean. Para ello, reunimos bibliografía relevante acerca de las nociones involucradas en el trabajo -CDIs, IA, MLL², etc.-, y sobre casos de uso de dicha herramienta llevados a cabo con distintas metodologías, utilizando Chat GPT para digerir toda esta información, ya sea a través de preguntas o pidiéndole ejemplos. Cabe destacar que el uso de su motor de búsqueda integrado para hallar fuentes de información fue muy útil de cara a la investigación.

Cabe recordar que el objetivo principal del trabajo consistía en testar si los modelos basados en PLN impulsados por IA se pueden utilizar para la detección de patrones no convencionales -también alineación, en segundo plano- entre los CDIs y el MC OCDE versión 2017 con un grado de precisión suficiente, ofreciendo una manera escalable y sistemática de analizar un conjunto más amplio de tratados fiscales u otros documentos legales con datos semiestructurados³, en consecuencia reduciendo errores y optimizando el tiempo y costes de revisión por expertos en la materia.

Al principio, tanto alumno como tutor consideramos la posibilidad de comparar los resultados producidos a través de tres diferentes herramientas:

- i. Por un lado, IAs especializadas en Legaltech⁴, siendo estas entrenadas con *datasets* jurídicos.
- ii. Por otro lado, a través de herramientas de análisis de texto que no impliquen el uso de una IA, por ejemplo, a través de Python.
- iii. Por último, a través de una IA generalista con un MLL especializado en PLN.

Sin embargo, debido al límite de 10000 palabras establecido por la Facultad para realizar dicho TFG resultaba imposible aplicar tal enfoque. Por ello, decidimos conjuntamente reformular el trabajo usando un modelo MLL de IA especializado en PLN, justificando la elección concreta.

Tras ello, pensé en comparar todos los CDIs publicados por el Ministerio de Hacienda a la vez para ver si la IA era capaz de hacer esta tarea masiva sin errores, pero tras seguir investigando acerca de los MLL vi que estos tienen una “ventana de contexto”⁵, esto es, tienen un número de *tokens* limitado en cada conversación que tengamos con la IA.

Según la empresa creadora del modelo, OpenAI, cada *token* supone $\frac{3}{4}$ de una palabra en inglés aproximadamente, lo que suponía que era también imposible comparar todos los CDIs a la vez, pues cada convenio tiene alrededor de 8000 *tokens*. Además, tampoco era viable realizar este experimento a gran escala porque supondría que el estudiante debía revisar todos los CDIs

para evaluar y validar los resultados de la IA, lo cual excedía el objetivo del TFG, tanto por razones de tiempo como de recursos disponibles.

Finalmente, tanto tutor como alumno valoraron la situación para fijar la orientación final del trabajo la cual consistió en incluir los CDIs de los socios comerciales más relevantes -hasta siete- justificando el ranking con magnitudes económicas publicadas por organismos públicos, lo cual supuso una gran ayuda al tener así un desarrollo más viable del TFG, como se mostrará a lo largo de esta publicación.

2.2. DESARROLLO DEL TFG

PLANTEAMIENTO DEL PROBLEMA Y DESAFÍOS INICIALES

Una vez configurado el objeto y alcance del trabajo, la cuestión principal era desarrollar una metodología lo suficientemente robusta para conseguir el mejor output posible.

Como los más iniciados ya conocen, no bastaba con realizar las preguntas a Chat GPT sin más, pues este tipo de IAs conversacionales requieren ir afinando nuestras peticiones, dándoles los mayores detalles y contexto posibles de la respuesta que queremos conseguir.

Antes de enfrentarnos a los MLL hay dos grandes problemas que enfrentar:

1. En primer lugar, la generalidad de los conjuntos de datos de preentrenamiento de los MLL, que incluyen textos de múltiples áreas del conocimiento, pero no necesariamente profundizan en un nicho específico -en nuestro caso la fiscalidad internacional-. Como consecuencia, es posible que no tengan suficiente información precisa en su entrenamiento y, por lo tanto, den respuestas imprecisas o superficiales.
2. En segundo lugar, su falta de comprensión de un texto desde la perspectiva de un campo concreto (Qiu & Jin 2024). Aquí el problema radica en que, incluso si el modelo ha visto textos relacionados con un área concreta, no necesariamente comprende el contexto y las relaciones profundas desde ese punto de vista. Por ejemplo, si le preguntamos cuál es la inversión más segura en tiempos de crisis, la IA nos podría decir que el oro es la mejor opción- Aunque puede ser cierto, nos proporciona esta respuesta sin considerar factores adicionales como la inflación, tasas de interés, la liquidez de la inversión o sin valorar otro tipo de activos.

Estos obstáculos pueden dar lugar a lo que ya se conoce comúnmente como “alucinaciones”. Es decir, situaciones en las que estos modelos generan información que, aunque gramatical y contextualmente puede ser coherente, es incorrecta o carece de fundamento de acuerdo con los datos reales. Así, una alucinación en un modelo utilizado para asesoramiento médico podría llevar a diagnósticos incorrectos o tratamientos inapropiados.

Para solucionar el primer problema se proporcionó al modelo los CDIs analizados y el MC de la OCDE 2017 como fuentes de información primarias, mientras que el segundo inconveniente se intentó solucionar a través de ingeniería del *prompt*, aumentando su comprensión contextual -véase la Figura 5-.

SELECCIÓN DE PAÍSES Y ALCANCE DEL ANÁLISIS

Por otro lado, cabe destacar cómo se realizó la selección de los países más relevantes para España desde el punto de vista económico, teniendo en cuenta los datos publicados por el Ministerio de Economía, Comercio y Empresa del año 2023 -los últimos datos disponibles en aquel momento- sobre exportaciones, importaciones y flujos de inversión directa, como podemos observar en la Tabla 1.

Tabla 1. Los países económicamente más relevantes para España en 2023 -datos provisionales-

Posición	País	Importaciones (M€)	Exportaciones (M€)	IED ⁷ en España (M€)	IEE ⁸ (M€)	Total (M€)
1	Francia	39.628,5	59.904,9	2.583	489	102.605,4
2	Alemania	47.057,9	40.018,5	3.000	1.055	91.131,4
3	Italia	28.301,6	33.044,8	246	171	61.763,4
4	Estados Unidos	28.267,8	18.904,2	8.146	5.850	61.168
5	China	44.244,2	7.578,7	131	92	51.822,9
6	Portugal	16.535,5	31.817,4	184	592	49.129,9
7	Reino Unido	10.796,8	22.564,2	3.693	726	37.780

Fuente: Ministerio de Economía, Comercio y Empresa (2024)

A continuación, elaboramos una metodología sólida orientada a alcanzar el objetivo del trabajo, soportada por las directrices mencionadas por OpenAI de cara a entrenar sus MLL y optimizar su funcionamiento, a la vez que la de autores como Nasser et al. (2023), De Silva et al. (2024), Cao et al. (2024), OpenAI (2024), etc.

ELECCIÓN DEL MODELO Y CREACIÓN DEL PROMPT

Un aspecto central de nuestro estudio fue la elección del modelo de inteligencia artificial más adecuado. Para ello, consideramos diversas métricas, entre las cuales destacamos las siguientes, por su especial relevancia para los fines del proyecto:

i. *Massive Multitask Language Understanding (MMLU)*: Evalúa el conocimiento general y la capacidad de razonamiento del modelo en una amplia variedad de temas, incluidos Derecho y Economía, aspectos altamente relevantes para este proyecto.

ii. *Discrete Reasoning Over Paragraphs (DROP)*: Evalúa la capacidad del modelo para extraer y razonar sobre datos factuales en documentos extensos.

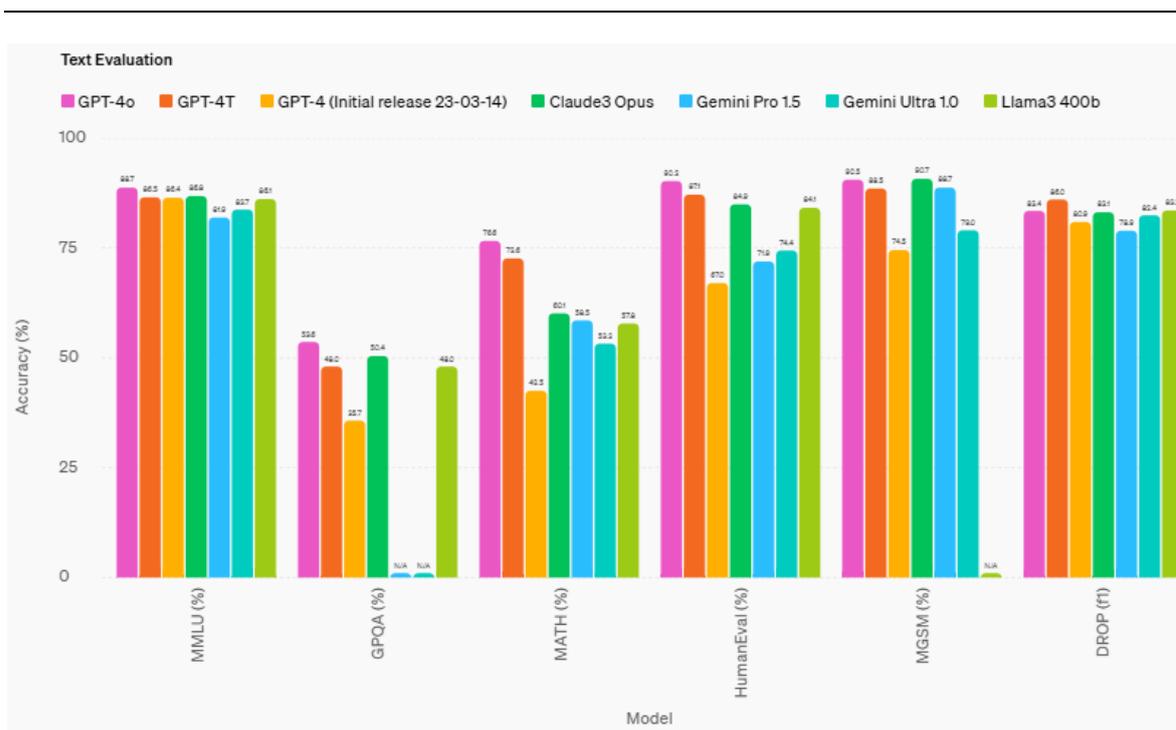
iii. *Graduate-Level Google-Proof Question Answering (GPOQA)*: Evalúa el razonamiento avanzado y el conocimiento específico del dominio de los modelos de IA. Mide el

rendimiento del modelo al responder preguntas abiertas en función de su comprensión del contexto y el razonamiento.

Tal como se muestra en la Figura 1, GPT-4o superaba a sus competidores en la gran mayoría de ellas, siendo solo superada por un escaso margen en una métrica llamada *DROP*.

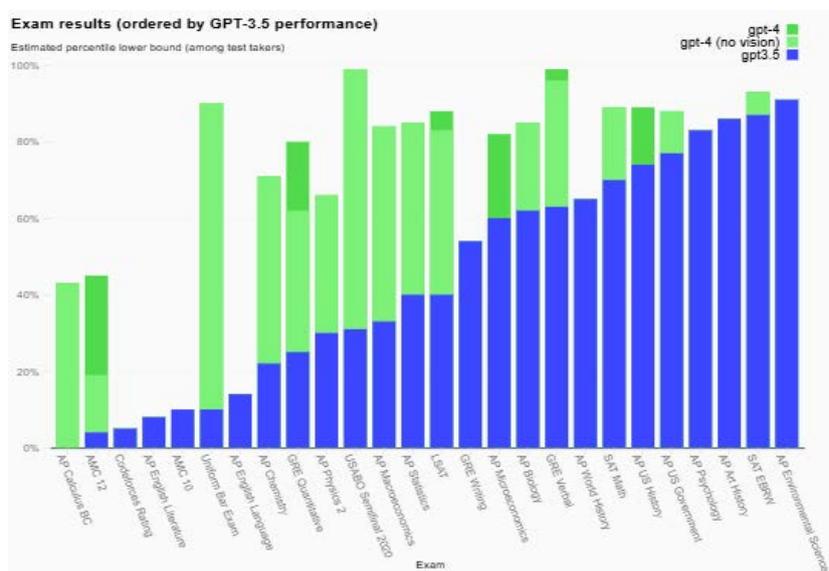
Por otro lado, ya que la IA tiene que tratar con textos legales como son los CDIs, también tuve en cuenta los resultados que obtuvo la versión anterior a GPT-4o, “GPT-4”, en el *Uniform Bar Examination (UBE)* (Katz, Bommarito, Gao, & Arredondo, 2024), una prueba estandarizada que evalúa las competencias de los aspirantes a abogados en Estados Unidos. Su objetivo es proporcionar una calificación uniforme que pueda ser utilizada para solicitar la admisión en múltiples jurisdicciones que han adoptado el *UBE*, situándose este modelo en el percentil 90 de los postulantes, como podemos observar en la Tabla 2.

Figura 1. Evaluación de textos de diferentes MLL



Fuente: OpenAI. (2024c).

Figura 2. GPT-4 vs GPT-3.5 resultados de exámenes simulados



Fuente: OpenAI. (2023)

Tabla 2. GPT-4 vs GPT-3.5 resultados de exámenes simulados en percentiles

Examen simulado EE. UU.	GPT-4 (percentil estimado)	GPT-3.5 (percentil estimado)
Examen Uniforme de Acceso a la Abogacía	298/400 (~90)	213/400 (~10)
Examen de Admisión a Escuelas de Derecho	163 (~88)	149 (~40)
Lectura y Escritura Basadas en Evidencia	710/800 (~93)	670/800 (~87)
Examen General de Acceso a Posgrados (GRE)	169/170 (~99)	154/170 (~63)

Fuente: OpenAI. (2023)

Las principales tareas que el modelo es capaz de desarrollar son síntesis, transformación y extracción de información. Este era el único modelo con capacidad de analizar archivos *.pdf*, *.doc*, etc., lo que es una cuestión relevante, pues le introdujimos los archivos *.pdf* de los CDIs y del MC OCDE versión 2017.

Respecto a subir archivos al modelo tuvimos en cuenta múltiples limitaciones, de espacio, tamaño, número de *tokens*, etc. Tras ello, decidimos que la mejor manera de hacer la comparación era en conversaciones separadas, es decir, haciendo una comparación por conversación, pues podría contaminarse el *output* de la IA al subir varios CDIs en la misma conversación o podría producir respuestas parciales por exceder los límites antes mencionados.

El idioma de los textos también es una variable a tener en cuenta cuando se trabaja con este tipo de herramientas. En el presente caso, escogimos hacer el TFG en inglés; sin embargo, no todos los CDIs estaban disponibles en inglés en la web del Ministerio de Hacienda.

Por ello, solo los convenios con Alemania y Reino Unido se descargaron en inglés; los demás estaban únicamente disponibles en español. No cabe duda de que lo óptimo hubiese sido que todos los CDIs estuviesen en el idioma que se ha hecho el trabajo y en el idioma en que se va a utilizar la IA -inglés- pero, por otro lado, esta disparidad de lenguaje ha permitido comprobar la capacidad de traducción de la herramienta.

He de añadir que el CDI con EE.UU. se ha unido con el protocolo que lo modifica en 2019, con el fin de ver si la herramienta es capaz de hacer una lectura comprensiva del documento teniendo en cuenta las disposiciones modificadas y no las derogadas.

Asimismo, tuvimos que comprobar que los textos adjuntados estuviesen escaneados en OCR⁹ para que la herramienta pudiese analizarlos al completo, además de renombrar los nombres de los archivos para que la tarea fuese lo más clara posible para el asistente., llamando a los archivos de la siguiente forma:

- “OCDE Model Tax Convention 2017”
- “Spain_Foreigncountry¹⁰_DTA.pdf”

Por último, el MC OCDE 2017 fue abreviado excluyendo partes no analizadas como la introducción, los comentarios sobre los artículos u otras disposiciones, de forma que el *.pdf* solo incluyese los artículos del convenio.

En resumen, si bien no ha sido posible conseguir una uniformidad absoluta entre los documentos adjuntados, con estas adaptaciones nos hemos aproximado lo máximo posible.

Otra parte clave del trabajo fue crear los *prompt* que íbamos a insertar al MLL. Para maximizar su eficiencia investigamos acerca de estrategias y métodos de ingeniería del *prompt*, tratando de reunir toda la información posible para la creación de un *prompt* introductorio que sería posteriormente insertado en todas las conversaciones. La guía de la API de OpenAI (2024g) menciona algunas estrategias para mejorar el resultado de nuestras indicaciones:

1. Redactar instrucciones claras: Como expone OpenAI: “Cuanto menos tenga que adivinar el modelo sobre lo que quieres, más probable será que lo consigas”. Las tácticas para lograr esta estrategia podrían ser:

- Incluir detalles y contexto en la indicación para obtener respuestas más relevantes, dejando el mínimo espacio posible para que el modelo adivine lo que quieres.
- Pedir al modelo que adopte una personalidad.

2. Proporcionar texto de referencia: Proporcionaremos al modelo información fiable y relevante para responder a las preguntas que le formularemos, evitando respuestas y citas falsas. Además, le indicaremos que responda con citas de los documentos subidos, haciendo referencia a artículos, disposiciones, páginas, etc. Esto facilita la futura revisión por parte de un supervisor humano, garantizando la trazabilidad y la verificabilidad.

3. Dividir las tareas complejas en subtareas más simples: Las tareas complejas tienden a tener tasas de error más altas que las tareas más simples, lo cual generalmente puede evitarse, ya que algunas tareas complejas pueden redefinirse como un flujo de trabajo de tareas más simples en el que los resultados de las tareas anteriores se utilizan para construir las entradas de las tareas posteriores.

Una táctica útil recomendada por OpenAI es resumir documentos largos por partes y construir un resumen completo recurrentemente. Esto podría hacerse fácilmente copiando y pegando solo el artículo que queremos comparar del DTA y el MC, pero esto sería difícil e ineficiente al comparar grandes cantidades de archivos, y una táctica contraria al propósito de este proyecto, que es lograr una detección automatizada.

4. Darle tiempo al modelo para que "piense", como lo hacen los seres humanos antes de dar una respuesta. Indicarle al modelo que calcule su propia solución a partir de principios básicos antes de apresurarse a llegar a una conclusión a veces implica mejores resultados. Otra táctica verdaderamente relevante para nuestro proyecto, relacionada con esta estrategia, es preguntar al modelo si ha omitido algo en pasadas anteriores. Como indica OpenAI: “Supongamos que usamos un modelo para listar extractos de una fuente relevantes para una pregunta específica. Después de listar cada extracto, el modelo debe determinar si debe comenzar a escribir otro o si debe detenerse. Si el documento fuente es extenso, es común que el modelo se detenga demasiado pronto y no enumere todos los extractos relevantes. En ese caso, a menudo se puede obtener un mejor rendimiento al solicitar al modelo consultas de seguimiento para encontrar los extractos que omitió en pasadas anteriores”.

Teniendo en cuenta las anteriores indicaciones, en la redacción del *prompt* introductorio nos ayudó parcialmente la propia IA, siendo necesario además introducir pequeñas modificaciones lingüísticas en el *output* obtenido a través de ella. Ejemplo del proceso de creación de la instrucción introductoria al modelo fue el siguiente:

- i. Primer intento: “Encuentra diferencias clave entre ambos documentos”.
- ii. Segundo intento: “Actúa como asesor fiscal español experto en Convenios de Doble Imposición firmados por España y especializado en el MC OCDE 2017”.
- iii. Tercer intento: “(...). Utiliza únicamente los documentos proporcionados. No recurras a conocimiento preentrenado salvo que se te indique expresamente”.
- iv. Cuarto intento: “(...). Analiza las disposiciones artículo por artículo o tema por tema, basándote en las preguntas que te proporcionaré más adelante”.

v. Quinto intento: “(...). Si la información parece poco clara, ambigua o falta, expón claramente la limitación y propón pasos para seguir investigando. Evita a toda costa las suposiciones”.

Los resultados iban mejorando a medida que íbamos acumulando las distintas tácticas y estrategias, hasta que se consideró que se llegó al punto óptimo, cuya instrucción final fue la siguiente:

“Vas a actuar como asesor fiscal español experto en Convenios de Doble Imposición (CDI) firmados por España y otros países, especializado en el marco jurídico establecido por la Organización para la Cooperación y el Desarrollo Económico (OCDE) en su versión Modelo de Convenio Tributario 2017. Su tarea consiste en asistir en un análisis jurídico comparativo entre el MC 2017 de la OCDE y el CDI firmado por España con [X país].

A) Documentos de referencia:

En primer lugar, subiré:

- 1. El MC 2017 de la OCDE como documento de referencia.*
- 2. El CDI firmado por España con [X país] para compararlos.*

Después de hacer esto no realizas ninguna acción hasta que te haga la primera petición relacionada con ellos.

B) Instrucciones para la tramitación y salida:

1. Espera hasta que se te solicite: No realice ninguna acción después de cargar los documentos hasta que le haga explícitamente una pregunta relacionada con el análisis.

2. Utiliza citas y estructura:

-Incluye citas que muestren los números exactos de los artículos y de las páginas de la información encontrada en ambos documentos.

3. Razona paso a paso:

-Analiza las disposiciones artículo por artículo o tema por tema basándote en las preguntas que le facilito.

4. Destaca los problemas y las lagunas posibles:

-Si la información parece poco clara, ambigua o falta, expón claramente la limitación y proponga pasos para seguir investigando, evite a toda costa las suposiciones. Compruebe dos veces los documentos para asegurarte de que no falta ningún fragmento.

Nota importante:

Debes centrarte únicamente en los documentos proporcionados y no basarte en conocimientos previos a menos que se te indique explícitamente que lo haga. Haz referencia a los archivos cargados como fuentes primarias”.

Posteriormente, elaboramos otros diez *prompt* que implicaban realizar una comparación de las disposiciones más relevantes de los CDIs -definición de establecimiento permanente, intereses, ganancias de capital, etc., como observamos en la Tabla 3-. Aquí también nos ayudamos de la IA para hacer preguntas más o menos extensas, con ligeras modificaciones para dar una mayor uniformidad y evitar redundancias.

Tabla 3. *Follow-up prompts*

P#	Complejidad	Materia	Instrucciones de seguimiento
P1	Baja	Ámbito de aplicación objetivo	Compare el ámbito de aplicación objetivo (artículo 2) del MC 2017 de la OCDE con el España_Italia_CDI (1980). ¿En qué difieren los tipos de impuestos cubiertos por los tratados? Destaque las desviaciones notables, si las hubiera.
P2	Alta	Establecimiento Permanente (EP)	Analice el artículo 5 sobre establecimientos permanentes en el MC 2017 de la OCDE y compárelo con las disposiciones relacionadas en el España_EE.UU._CDI (2019). ¿Qué diferencias existen en la definición del concepto?
P3	Baja	Beneficios empresariales	Evalúe el artículo 7 del MC 2017 de la OCDE sobre beneficios empresariales y las disposiciones relacionadas en España_Alemania_CDI (2012). ¿Existen desviaciones?

P4	Media	Rentas de trabajo	Analice el artículo 15 del MC 2017 de la OCDE sobre los rendimientos del trabajo y compárelo con las disposiciones correspondientes del España_Francia_CDI (1997). ¿Cómo abordan estos tratados la tributación de los sueldos, salarios y otras remuneraciones similares percibidas por personas físicas que trabajan en el otro Estado contratante? ¿Existen diferencias en cuanto a exenciones, umbrales temporales (norma de los 183 días) o condiciones específicas para la tributación en el Estado de la Fuente? Destaque las desviaciones notables, si las hubiera.
P5	Alta	Intereses	Compare las disposiciones del artículo 11 del MC OCDE 2017 con las del España_EE.UU._CDI (2019) relativas a la tributación de los intereses. Destaque las desviaciones notables, si las hubiera.
P6	Baja	Cánones	Compare el artículo 12 del MC 2017 de la OCDE sobre cánones con las disposiciones del España_Alemania_CDI. ¿Cómo abordan estos tratados la fiscalidad de los cánones, incluidos los tipos de retención a cuenta, las definiciones de cánones y las exenciones para derechos de propiedad intelectual específicos? Destaque las diferencias notables, si las hubiera.
P7	Baja	Rentas inmobiliarias	Examine el artículo 6 del MC 2017 de la OCDE sobre rentas de la propiedad. ¿En qué difieren o se alinean las disposiciones del España_ReinoUnido_CDI (2014) al abordar este concepto?

P8	Alta	Ganancias de capital	Analice el artículo 13 del MC 2017 de la OCDE sobre plusvalías y compárelo con las disposiciones del España_Portugal_CDI (1995). ¿Cómo abordan estos tratados la tributación de las ganancias derivadas de acciones y activos mobiliarios? Discuta si el España_Portugal_CDI incorpora excepciones o disposiciones especiales para las sociedades patrimoniales, las transmisiones indirectas o la participación sustancial. ¿Existen cláusulas antiabuso específicas para evitar el treaty shopping o las transferencias artificiales? Destaque cualquier desviación en los métodos de cálculo o en las normas relacionadas con la valoración y el calendario de las ganancias.
P9	Alta	Eliminación de la doble imposición	Analizar el artículo 23 del MC 2017 de la OCDE, que establece métodos para eliminar la doble imposición (métodos de exención y de crédito). ¿Está el España_China_CDI (2021) alineado con estos métodos o introduce mecanismos alternativos? Destaque las desviaciones notables, si las hubiera.
P10	Media	Provisiones de no discriminación	Evalúe el artículo 24 sobre no discriminación en el MC de la OCDE de 2017 y en el España_Francia_CDI. ¿Cómo protegen estos tratados contra la desigualdad de trato fiscal a los nacionales o residentes de los Estados contratantes?

Fuente: Elaboración propia

SISTEMA DE EVALUACIÓN Y VALIDACIÓN DEL OUTPUT DE LA IA

En aquellos casos que la IA tenía fallos en la respuesta, ya sea a través de omisiones, alucinaciones o por malas interpretaciones, se ha llevado a cabo correcciones iterativas insertando nuevos *prompt*, de manera no demasiado exhaustiva para ver si la herramienta era capaz de corregir sus respuestas, con un máximo de tres intentos al respecto, ya que estamos valorando la capacidad de la IA para acertar de su primera respuesta, y en casos excepcionales corregirla.

Este enfoque respondía a la necesidad de que la herramienta funcione de manera eficiente y autónoma, sin requerir una supervisión constante por parte de un experto en la materia, ya que un proceso de revisión manual excesivo limitaría su aplicabilidad práctica y reduciría su utilidad en entornos profesionales y académicos.

En cuanto al sistema de evaluación tomé como referencia el de Contreras (2024), utilizando distintos criterios y asignándoles un peso diferente en función de su relevancia. Cada criterio se ha evaluado acorde a una escala de 5 puntos desde deficiente (1) a excelente (5), idea dada por la IA y tomada en consideración al parecernos un buen modo de evaluación de cada criterio, como podemos observar en la Tabla 4.

La asignación de pesos ponderados en la Tabla 4 respondía a la importancia relativa de cada criterio respecto al objetivo del TFG. Se ha otorgado el mayor peso al análisis comparativo (35%), pues era clave que la IA fuese capaz de identificar y razonar las similitudes y diferencias entre los CDIs y el Modelo de Convenio OCDE. Le sigue la precisión de las referencias (30%), dado que la exactitud de las citas extraídas de los documentos era esencial para asegurar una mayor trazabilidad por parte del experto o supervisor, especialmente ante el riesgo de "alucinaciones" propias de los modelos de IA.

La relevancia (25%) es un criterio que también se valora significativamente, al ser clave que las respuestas se ajusten al contenido y contexto del *prompt* introducido, aunque se considera menos determinante que los aspectos anteriores. Finalmente, la exhaustividad (10%) recibe un peso menor, ya que, si bien es deseable una cobertura completa, en este trabajo se ha priorizado la precisión y profundidad frente a la cantidad de contenido generado, tratando de omitir detalles menores e información no esencial.

Tabla 4. Método de evaluación de los resultados de la IA

Criterio	Puntuación 5 (Excelente)	Puntuación 3 (Aceptable)	Puntuación 1 (Deficiente)
Relevancia (25%)	Responde completamente al <i>prompt</i> y al contexto proporcionado.	Responde parcialmente a la pregunta, pero omite algunos aspectos o incluye detalles irrelevantes.	La respuesta es irrelevante o se desvía completamente del tema.
Precisión de las referencias (30%)	Proporciona citas correctas y citas textuales extraídas de los archivos fuente.	Incluye algunas citas incorrectas o referencias erróneas, requiriendo verificación manual.	Contiene referencias falsas o citas inventadas que comprometen la fiabilidad.
Análisis comparativo (35%)	Identifica desviaciones clave, similitudes e implicaciones con	Señala diferencias, pero carece de profundidad en los análisis o ignora implicaciones legales.	Presenta observaciones superficiales sin un análisis adecuado o ejemplos.

Exhaustividad (10%)	razonamiento bien fundamentado.	
	Cubre todos los aspectos del <i>prompt</i> de manera completa, sin omisiones.	Aborda los puntos principales, pero omite algunos detalles menores o excepciones. Deja elementos clave sin abordar, lo que hace que la respuesta sea incompleta

Fuente: Elaboración propia

Finalmente, las ecuaciones para evaluar los resultados de la IA fueron las siguientes:

Ecuación 1:

$$WS_q = (w_1 \cdot SC_1) + (w_2 \cdot SC_2) + (w_3 \cdot SC_3) + (w_4 \cdot SC_4) \quad [1]$$

Ecuación 2:

$$WS_{total} = \frac{\sum_{q=1}^{10} WS_q}{10} \quad [2]$$

Donde:

- WS_q : Puntuación ponderada por pregunta
- w_1, w_2, w_3, w_4 : Pesos asignados a los cuatro criterios de evaluación.
- SC_1, SC_2, SC_3, SC_4 : Puntuaciones (1–5) otorgadas para cada criterio respectivo.
- WS_{total} : Puntuación total ponderada

$$\text{Objetivo: } WS_{total} \geq 3.5$$

Tras haber configurado la metodología anteriormente mencionada, procedimos a llevar a cabo los distintos pasos, generando la herramienta un *output* en cada respuesta el cual evaluamos acorde a dichos criterios. Para contrastar la información revisamos cada disposición comparada entre el correspondiente CDI y el MC OCDE 2017, es decir, una revisión manual párrafo a párrafo para discernir las diferencias y similitudes entre ambos textos legales¹¹.

Es posible que este método de evaluación sea considerado considerablemente subjetivo, pues implicaba que un alumno evalúe por sí mismo el desempeño de la IA en la consecución del fin del trabajo, por lo que en aquellos casos en los que se estime oportuno es posible que se añada la valoración adicional del director del TFG, o al menos su revisión. En el presente caso el alumno adjuntó como anexos los *outputs* producidos por el *chatbot*, para que estos fuesen posteriormente revisados tanto por su tutor como por el tribunal evaluador, con el fin de mostrar la mayor transparencia posible en su proceso de elaboración.

2.3. RESULTADOS

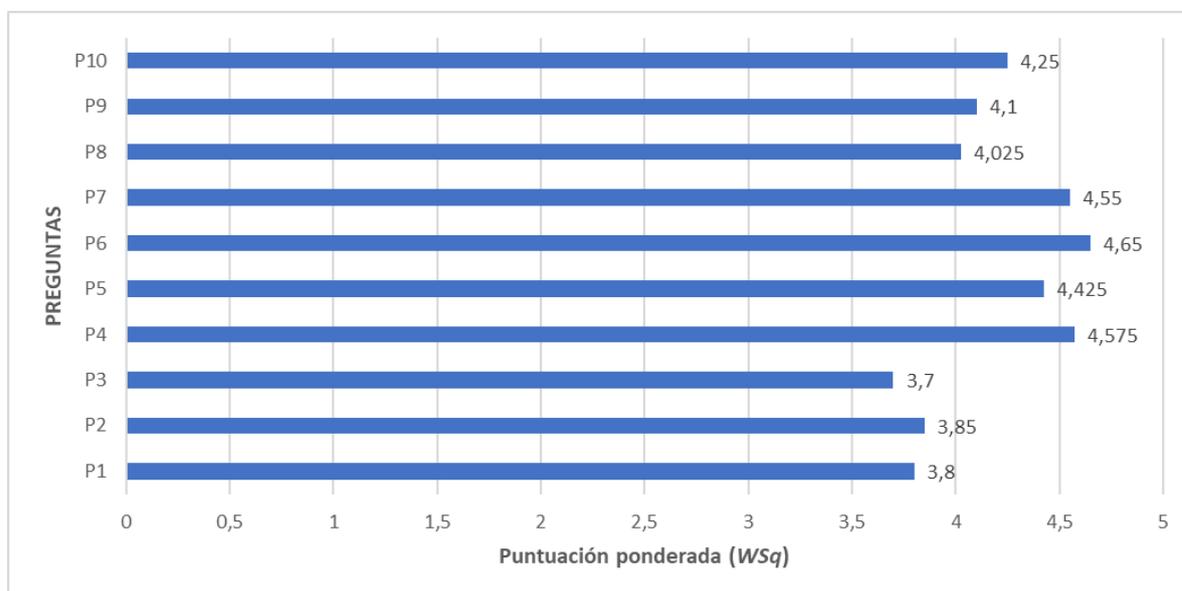
En esta sección se resume de forma sintética los resultados obtenidos en el TFG, de cara a ilustrar la calidad de la respuesta generada con este tipo de herramientas.

En primer lugar, cabe destacar que, aunque el MLL logró una puntuación media ponderada de 4.19/5, se requirió de verificación para corregir algunos errores. En la Figura 3 podemos observar que las preguntas Q4, Q6 y Q7 obtuvieron puntuaciones ponderadas superiores a 4,5, siendo la Q6 la que casi alcanzó una puntuación perfecta (4,65), seguida de la Q5 (4,55). Por otro lado, las preguntas Q1, Q2, Q3, Q8 y Q9 mostraron un rendimiento inferior, causado por imprecisiones o insuficiente profundidad comparativa. La Q3 obtuvo la peor puntuación media ponderada (3,7) debido a la omisión de detalles críticos.

El criterio de relevancia (4,9) fue muy consistente en todas las preguntas, mientras que el análisis comparativo (4,05) mostró un resultado satisfactorio para la mayoría de las preguntas, excepto para la pregunta 8. Podría mejorarse preguntando explícitamente las implicaciones fiscales de las desviaciones y alineaciones en la pregunta introductoria.

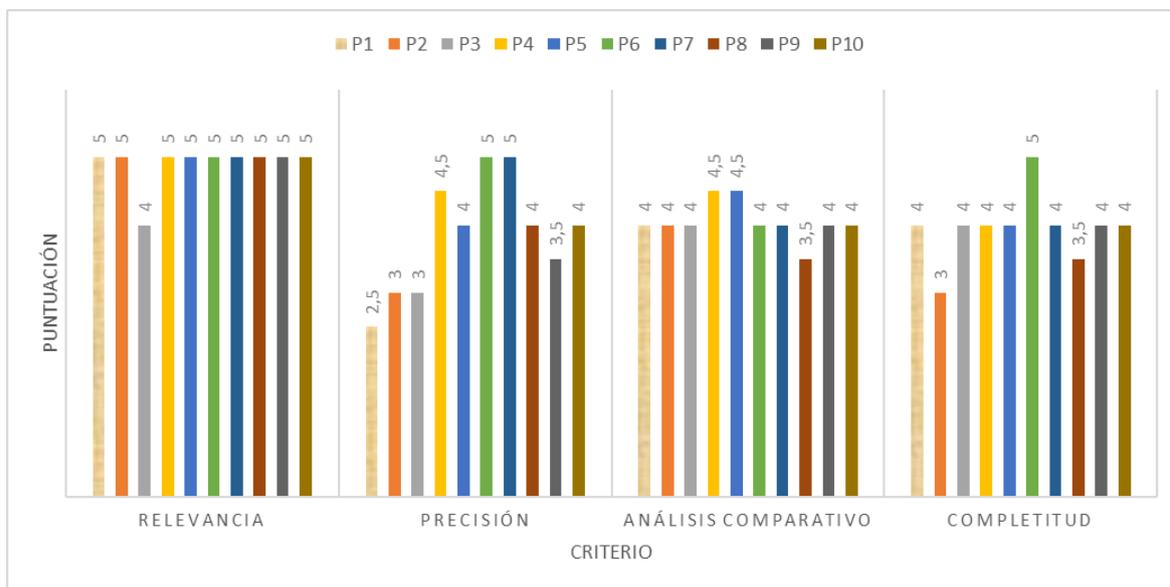
El criterio de precisión (3,85) obtuvo la puntuación media más baja, siendo el criterio con la mayor variabilidad como consecuencia de alucinaciones u omisiones.

Figura 3. Puntuación Ponderada por Pregunta (SC)



Fuente: Elaboración propia

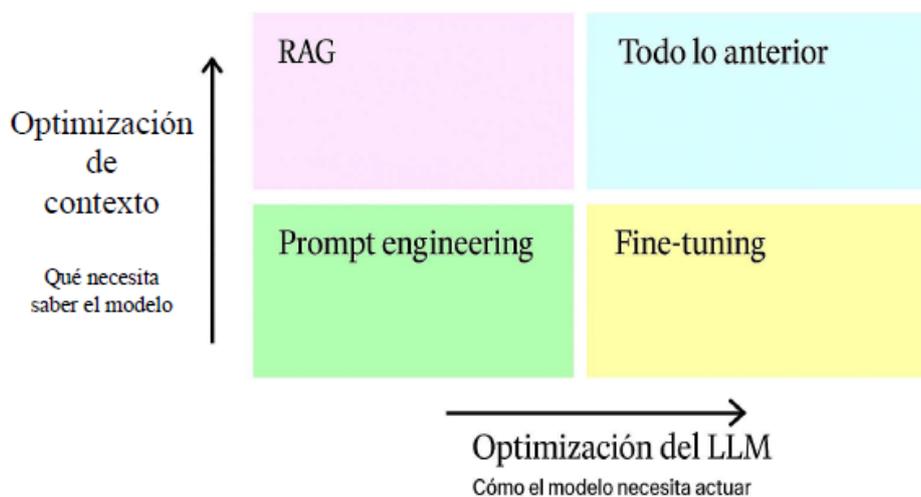
Figura 4. Puntuación Media por criterio (SWq)



Fuente: Elaboración propia

Como podemos observar en la Figura 4, hemos conseguido resultados considerables solo utilizando ingeniería del *prompt* y carga manual de documentos *.pdf* teniendo margen de mejora en futuras aplicaciones si se empleasen otras estrategias más complejas no consideradas por su dificultad técnica, como podemos ver en la Figura 5.

Figura 5. Cadena de optimización MLL



Fuente: OpenAI (2023)

3. CONCLUSIONES

Como conclusión general de la elaboración del TFG con la asistencia de la herramienta, cabe destacar que, si bien la IA ha servido como asistente a lo largo del proceso, la investigación fue llevada a cabo de modo habitual en lo que concierne a bibliografía, casos de uso similares, búsqueda de datos económicos sobre países más relevantes, evaluación del *output* y conclusiones del trabajo. Eso sí, la herramienta facilitó la obtención de información sobre distintos conceptos -tanto fiscales como de su ámbito- y a fundamentar las decisiones que hemos ido tomando a lo largo del trabajo aportando solidez y robustez respecto a las magnitudes económicas escogidas para hacer el ranking de países, la construcción del *prompt* introductorio, la elección del MLL, etc.

A pesar de que este trabajo es un experimento a pequeña escala, la herramienta ha producido un *output* lo suficientemente preciso como para ser tomado seriamente en cuenta en futuras investigaciones de mayor dimensión, no solo de fiscalidad internacional, sino en cualquier rama del Derecho y de la Economía y la Empresa.

La estrategia combinada de contextualización con documentos específicos, ingeniería de *prompt* y validación iterativa permitió adaptar un MLL generalista -GPT-4o- a un dominio altamente técnico/específico. Esto demuestra que, con un diseño metodológico riguroso, los MLL pueden convertirse en herramientas auxiliares eficaces para análisis legales y económicos complejos, aunque su implementación requiere supervisión experta para garantizar su fiabilidad.

Por ello, podemos concluir que la IA, siempre y cuando se utilice de forma precavida contrastando sus respuestas, no se trata de una herramienta que sustituya de forma plena el trabajo que debe llevar a cabo tanto alumnos como expertos en una determinada materia, sino supone un complemento multiplicador de los resultados que obtienen en función de su conocimiento y las horas trabajadas, permitiéndoles llegar más lejos que lo que sus limitaciones humanas hacen posible.

Incluso, parece razonable pensar que estas herramientas son mucho más eficientes y útiles para aquellos/as que más conocimiento tengan de la materia tratada, ya que podrán evaluar la calidad de su respuesta, hacer mejores preguntas y tener un pensamiento crítico de lo que genere el asistente, mientras que alguien inexperto puede tomar como correctas sus “alucinaciones”.

En resumen, con una orientación básica, la herramienta ha permitido desarrollar este trabajo con recursos limitados, ya sea en términos de tiempo, conocimiento o extensión del Trabajo de Fin de Grado, lo que ha facilitado la comprensión de conceptos ajenos al ámbito de estudio del alumno de manera rápida y sencilla.

Además, ha permitido aplicar estos conocimientos de forma práctica, demostrando que este enfoque es escalable y posee un gran potencial para futuras investigaciones, pues aún queda por probar múltiples estrategias y metodologías no utilizadas en este trabajo, como: (i) mejorar *Retrieval-Augmented Generation (RAG)*; (ii) *fine-tuning* del modelo; (iii) combinación de todas las estrategias mencionadas, como podemos ver en la Figura 5.

Para cerrar el apartado de conclusiones, se recogen a continuación algunas reflexiones que emergen de la realización del TFG. Serían las siguientes:

Primero, desde el lanzamiento de Chat GPT en noviembre de 2022, basado en el modelo GPT-3.5, con gran interés y curiosidad el autor ha utilizado estas herramientas para

complementar su formación con la capacidad de estas tecnologías, pues desde un primer momento percibió que bien entrenadas, podían servir como tutor personal con disponibilidad total para todas y cada una de las asignaturas de su programa de estudios.

Segundo, es cierto que, debido a que su base de datos no está totalmente actualizada, este tipo de herramientas rara vez han sido utilizadas por el alumno para el estudio de las diferentes ramas de Derecho, pues es crucial tener tanto legislación como jurisprudencia actualizada en este ámbito de las ciencias sociales, constantemente sometida a cambios -Aunque actualmente nuevas características introducidas a comienzo de 2025 como la denominada “investigación profunda” ha resultado increíblemente útil para llevar a cabo investigaciones jurídicas-.

Además, consideraba que al estar la IA de OpenAI pre-entrenada con internet como base de datos no era lo suficientemente precisa y fiable como para investigar sobre este campo, pues el Derecho es particular de cada país, y cada país lo configura de manera interna -salvo el de carácter internacional-. Sin embargo, lo contrario sucede en cuanto a la Economía, Finanzas y la Administración de Empresas, pues los criterios y conceptos utilizados siguen una mayor uniformidad a nivel mundial en comparación al Derecho.

Tercero, hay que tener en cuenta que la mayor parte de investigación, bibliografía y casos prácticos entorno a estos campos se ha desarrollado en lengua inglesa, lo cual ha facilitado sumamente adaptarme al uso de la IA para explicar y dar ejemplos de todo aquello que no entendía de estos ámbitos, al realizar el Grado en Administración de Empresas en inglés.

Cuarto, el autor siempre ha tenido más dificultad para entender fórmulas, gráficos y realizar casos prácticos numéricos. Así, mientras que los primeros años de carrera experimentó dificultades con asignaturas como Macroeconomía, Microeconomía o Matemáticas Empresariales, en las asignaturas más vinculadas al área de Derecho o iba con total fluidez y facilidad. Por ello, desde la introducción de ChatGPT, teniendo en cuenta la experiencia del autor como estudiante universitario, se considera que el uso del *chatbot* como tutor ha permitido mejorar de forma significativa no sólo sus resultados académicos sino, más importante aún, la comprensión real sobre estos campos.

Quinto, un tema recurrente cuando se habla de la IA en la educación es el miedo que tienen los/as docentes en cuanto al plagio y la falta de razonamiento y trabajo por parte de los alumnos. Este inconveniente creemos que se puede solucionar, ya que solo basta ver si en un examen final de una determinada asignatura es capaz el alumno de aprobarlo y sacar nota en él, dado que ninguna herramienta de este tipo se puede utilizar cuando el alumno se enfrenta al “papel y bolígrafo” de toda la vida, o cuando este realiza una exposición oral de sus conocimientos frente al profesor o a la clase.

Sexto, en el caso concreto del TFG, el papel del tutor ha seguido siendo crucial para realizarlo, pues ha seguido orientando tanto en la toma de decisiones metodológicas clave como en el enfoque del trabajo, mientras que la IA ha permitido solucionar errores menores y automatizar trabajos de carácter más mecánico y secundario -aunque es cierto que a la hora de detectar patrones no convencionales ha sido una pieza fundamental, como se ha observado-.

Como consecuencia, se consigue que el rol del tutor se centre donde de verás genera un verdadero valor añadido, y no tanto en detalles de poca envergadura, ya que el tutor experto especializado en el ámbito objeto de estudio del TFG, sabe delimitar mejor las cuestiones de fondo y de formato que el alumno ayudado de una IA, pues, como se ha expuesto anteriormente, la IA actúa como un multiplicador del conocimiento que uno mismo tiene.

En resumen, la experiencia pone de manifiesto como la IA ha afectado a distintos campos de las ciencias sociales y cómo, con un uso ético y robusto, se puede conseguir resultados nunca vistos en la historia de la humanidad, en un increíblemente breve período de tiempo.

Se desconoce aún cuál es el límite en cuanto al desarrollo y aplicación de estas nuevas tecnologías, aunque, como todos los grandes inventos, se requiere de un uso ético y responsable de ellos para que prevalezca su impacto positivo sobre los efectos colaterales no deseados.

Notas

¹ El Procesamiento del Lenguaje Natural (PLN) es una rama de la inteligencia artificial que se centra en la interacción entre las computadoras y el lenguaje humano, permitiendo que las máquinas comprendan, interpreten y generen texto de manera similar a como lo hacen las personas.

² Un Modelo Largo de Lenguaje (MLL) es un tipo de inteligencia artificial diseñado para comprender, procesar y generar texto en lenguaje natural. Estos modelos están basados en arquitecturas avanzadas de aprendizaje profundo, como Transformers, y han sido entrenados con enormes cantidades de datos textuales.

³ Decimos que los datos son semiestructurados porque los documentos están estructurados en secciones, artículos, disposiciones finales, etc., pero la redacción y el formato difieren sustancialmente entre ellos.

⁴ Por ejemplo, actualmente el alumno se halla trabajando en un Departamento Jurídico-Tributario, donde ha tenido la oportunidad de probar diferentes herramientas: Maite AI, Vincent AI (Vlex), GenIA-L (Lefebvre), etc.

⁵ Una ventana de contexto más amplia permite al modelo mantener conversaciones más coherentes y analizar documentos más extensos sin perder información relevante. Sin embargo, ampliar esta ventana también implica mayores requerimientos computacionales y puede introducir desafíos, como la gestión de información irrelevante o redundante. Por ejemplo, el modelo GPT-4o tiene una ventana de contexto de 128.000 *tokens*, teniendo el *output* un límite de 16384 *tokens*.

⁶ La ingeniería de *prompts* es una disciplina emergente en el campo de la inteligencia artificial que se centra en el diseño y optimización de instrucciones, conocidas como "*prompts*", para guiar de manera efectiva a los modelos de lenguaje en la generación de respuestas deseadas.

⁷ Inversión Extranjera Directa (IED)

⁸ Inversión Española en el Extranjero (IEE)

⁹ Tecnología que permite convertir imágenes de texto mecanografiado, manuscrito o impreso en datos de texto que pueden ser procesados por una máquina

¹⁰ Por ejemplo, en el caso de China sería "Spain_China_DTA.pdf".

¹¹ Esta tarea evidenció la imposibilidad de comparar todos los CDIs a la vez, como inicialmente se quería hacer, pues revisar estas diez disposiciones conllevó un tiempo considerable.

Agradecimientos

El autor desea agradecer a su tutor del Trabajo de Fin de Grado y profesor de las asignaturas de Tax System I y II, Don Antonio Jesús Sánchez-Fuentes, la oportunidad de realizar dicho trabajo de una manera innovadora y disruptiva, la cual ha permitido al alumno adquirir conocimientos transversales acerca del uso de inteligencia artificial, PLN y MLL, campos ajenos a mi área de estudio, pero que cada vez tienen más incidencia en él, motivando al autor a seguir estudiando acerca de estos temas muy de cerca para el resto de mi carrera profesional.

Cabe destacar que, el primer día de clase que tuvimos con el tutor -profesor por aquel entonces-, él mencionaba en su presentación una famosa frase de Galileo Galilei, "No se puede enseñar nada a un hombre; solo se le puede ayudar a encontrar la respuesta dentro de sí mismo", frase que como se ha observado, ha llevado a cabo con éxito en mi caso personal al permitirme

explorar, cuestionar y construir mi propio conocimiento en lugar de simplemente imponerlo, siendo un ejemplo de la función del docente como guía del alumno para que desarrolle su propio potencial. Por todo ello, le estoy muy agradecido.

REFERENCIAS

- Cao, M., Wang, Q., Zhang, X., Lang, Z., Qiu, J., Yung, P. S. H., & Ong, M. T. Y. (2024). Large language models' performances regarding common patient questions about osteoarthritis: A comparative analysis of ChatGPT-3.5, ChatGPT-4.0, and Perplexity. *Journal of Sport and Health Science*. <https://doi.org/10.1016/j.jshs.2024.101016>
- Contreras, C. (2024). ¿Qué nota obtiene ChatGPT en un examen de economía pública? *e-pública: Revista electrónica sobre la enseñanza de la Economía Pública*, 35, pp. 42–75. <https://e-publica.unizar.es/es/articulo/que-nota-obtiene-chatgpt-en-un-examen-de-economia-publica-01>
- De Silva, A., Wijekoon, J. L., Liyanarachchi, R., Panchendrarajan, R., & Rajapaksha, W. (2024). AI insights: A case study on utilizing ChatGPT intelligence for research paper analysis. *Proceedings of the 14th International Workshop on Bibliometric-Enhanced Information Retrieval*. <https://doi.org/10.48550/arXiv.2403.03293>
- Katz, D. M., Bommarito, M. J., Gao, S., & Arredondo, P. (2024). *GPT-4 passes the bar exam*. *Philosophical Transactions of the Royal Society A*, 382. <https://doi.org/10.2139/ssrn.4389233>
- Ministerio de Economía, Comercio y Empresa (2024). *El sector exterior en 2023. Información Comercial Española: Boletín Económico* [Número especial], (3171), pp. 105-109, 141-151, 311-314, 355-359. <https://revistasice.com/index.php/SICE/issue/view/839>
- Ministerio de Hacienda. (1980). *Instrumento de Ratificación del Convenio entre España e Italia para evitar la doble imposición en materia de impuestos sobre la renta y para prevenir la evasión fiscal*. https://www.hacienda.gob.es/Documentacion/Publico/NormativaDoctrina/Tributaria/CDI/BOE_Italia.pdf
- Ministerio de Hacienda. (1990). *Instrumento de Ratificación del Convenio entre el Reino de España y los Estados Unidos de América para evitar la doble imposición y prevenir la evasión fiscal respecto de los impuestos sobre la renta*. https://www.hacienda.gob.es/Documentacion/Publico/NormativaDoctrina/Tributaria/CDI/BOE_EEUU.pdf
- Ministerio de Hacienda. (1995). *Instrumento de Ratificación del Convenio entre el Reino de España y la República Portuguesa para evitar la doble imposición y prevenir la evasión fiscal en materia de impuestos sobre la renta*. https://www.hacienda.gob.es/Documentacion/Publico/NormativaDoctrina/Tributaria/CDI/BOE_Portugal.pdf
- Ministerio de Hacienda. (1997). *Convenio entre el Reino de España y la República Francesa a fin de evitar la doble imposición y de prevenir la evasión y el fraude fiscal en materia de impuestos sobre la renta y sobre el patrimonio*. https://www.hacienda.gob.es/Documentacion/Publico/NormativaDoctrina/Tributaria/CDI/BOE_Francia.pdf

- Ministerio de Hacienda (2019). *Protocolo y su Memorando de entendimiento, hechos en Madrid el 14 de enero de 2013, que modifican el Convenio entre el Reino de España y los Estados Unidos de América para evitar la doble imposición y prevenir la evasión fiscal respecto de los impuestos sobre la renta, y su Protocolo*. <https://www.boe.es/boe/dias/2019/10/23/pdfs/BOE-A-2019-15166.pdf>
- Ministerio de Hacienda. (2021). *Convenio entre el Reino de España y la República Popular China para eliminar la doble imposición en relación con los impuestos sobre la renta y prevenir la elusión y evasión fiscales y su Protocolo*. <https://www.boe.es/boe/dias/2021/03/30/pdfs/BOE-A-2021-4911.pdf>
- Nasseri, M., Brandtner, P., Zimmermann, R., Falatouri, T., Darbianian, F., & Obinwanne, T. (2023). Applications of large language models (LLMs) in business analytics: Exemplary use cases in data preparation tasks. *Lecture Notes in Computer Science*. Pp. 182-198. https://doi.org/10.1007/978-3-031-48057-7_12
- OECD. (2017). *Model tax convention on income and on capital: Condensed version 2017*. OECD Publishing, Paris. https://doi.org/10.1787/mtc_cond-2017-en
- OpenAI. (2023). *GPT-4 research*. <https://openai.com/index/gpt-4-research/>
- OpenAI. (2024a). *Distillation*. OpenAI Platform. Retrieved December 11, 2024, from <https://platform.openai.com/docs/guides/distillation>
- OpenAI. (2024b). *File uploads FAQ*. OpenAI Help Center. Retrieved December 9, 2024, from <https://help.openai.com/en/articles/8555545-file-uploads-faq>
- OpenAI. (2024c). *Hello GPT-4o*. <https://openai.com/index/hello-gpt-4o/>
- OpenAI. (2024d). *Model selection*. OpenAI Platform. Retrieved December 9, 2024, from <https://platform.openai.com/docs/guides/model-selection>
- OpenAI. (2024e). *Models*. OpenAI Platform. Retrieved December 9, 2024, from <https://platform.openai.com/docs/models>
- OpenAI. (2024f). *Optimizing LLM accuracy*. OpenAI. Retrieved December 12, 2024, from <https://platform.openai.com/docs/guides/optimizing-llm-accuracy>
- OpenAI. (2024g). *Prompt engineering*. OpenAI Platform. Retrieved December 10, 2024, from <https://platform.openai.com/docs/guides/prompt-engineering>
- OpenAI. (2024h). *Text generation*. OpenAI Platform. Retrieved December 11, 2024, from <https://platform.openai.com/docs/guides/text-generation>
- OpenAI. (2024i). *Tokenizer*. OpenAI Platform. Retrieved December 11, 2024, from <https://platform.openai.com/tokenizer>
- Qiu, Y., & Jin, Y. (2024). ChatGPT and finetuned BERT: A comparative study for developing intelligent design support systems. *Intelligent Systems with Applications*, 21, 200308.
- Spanish Finance Ministry. (2012). *Agreement between the Kingdom of Spain and the Federal Republic of Germany for the avoidance of double taxation and the prevention of fiscal evasion with respect to taxes on*

income and on capital.
https://www.hacienda.gob.es/Documentacion/Publico/NormativaDoctrina/Tributaria/CDI/BOEIN_Alemania_2012.pdf

Spanish Finance Ministry. (2014). *Convention between the Kingdom of Spain and the United Kingdom of Great Britain and Northern Ireland for the avoidance of double taxation and the prevention of fiscal evasion with respect to taxes on income and on capital.*
https://www.hacienda.gob.es/Documentacion/Publico/NormativaDoctrina/Tributaria/CDI/BOEIN_RUnido.pdf

Assessing AI as a tool during the Final Degree Project on the field of Taxation: a complement or substitute for the student?

Abstract

This article presents the experience of a Law-Business Administration student in the development of a Final Degree Project (TFG) through artificial intelligence (AI). The research, within the field of international taxation, focused on identifying key differences in Double Taxation Agreements (DTAs) between Spain and seven countries, comparing them with the OECD Model Tax Convention (OECD MTC) of 2017. The article highlights how AI complements, but does not replace, the work of the student, who played a crucial role in validating and supervising the results generated by the tool. The student's point of view is used so that my colleagues can assess its use in future works/projects. Also, so that university professors can complement their vision when facing similar documents received from their students.

Keywords: artificial intelligence, natural language processing, double taxation agreements, international taxation, prompt engineering, final degree project, university teaching.

JEL codes: H25, H26, C63.