



# EL VIAJE DE ULISES: HACIA UN MODELO DE INTELIGENCIA ARTIFICIAL EXPLICABLE PARA RESOLVER PROBLEMAS DE LIBRE COMPETENCIA

Piero Alexis Malca Vilchez, Enzo Rodrigo Gomez Rojas y César Humberto Quiñones Costa

# El viaje de Ulises: Hacia un modelo de inteligencia artificial explicable para resolver problemas de libre competencia

Marzo 2025



## Piero Alexis Malca Vilchez

Bachiller en Derecho por la Pontificia Universidad Católica del Perú. Asociado encargado del Área de Innovación y Tecnología en Diez Canseco Abogados. ORCID: <https://orcid.org/0000-0001-8267-8568>



## Enzo Rodrigo Gomez Rojas

Abogado por la Pontificia Universidad Católica del Perú. Asociado del Área de Competencia en Miranda & Amado. Adjunto de docencia de los cursos de Instituciones de Derecho Mercantil y Derecho de la Competencia Desleal en la Facultad de Derecho PUCP. ORCID: <https://orcid.org/0009-0004-9000-401X>



## César Humberto Quiñones Costa

Abogado por la Universidad de Lima. Consultor de Cumplimiento en Libre Competencia en Pacifico EPS. Adjunto de docencia de Análisis Económico del Derecho en la Facultad de Derecho PUCP. ORCID: <https://orcid.org/0000-0001-8488-4218>

**Abstract:** This research contributes to the field of Explainable Artificial Intelligence (XAI) applied to the legal sector. A comprehensive literature review revealed a predominance of quantitative approaches to explainability, highlighting a gap in methods applicable to disciplines such as Law, where numerical methods face limitations. To address this need, we implemented and validated alternative non-quantitative approaches, such as Chain of Thought and Retrieval-Augmented Generation. These methods proved effective in enhancing the transparency of AI reasoning, enabling better tracking of information processing and response generation through formats accessible in commercial software. You can find our repository at the following link: [https://github.com/pmalca/ulises\\_project](https://github.com/pmalca/ulises_project)

**Resumen:** Esta investigación aporta al campo de la Inteligencia Artificial Explicable (XAI) aplicada al sector legal. Una revisión exhaustiva de la literatura destacó una predominancia de enfoques cuantitativos en la explicabilidad, evidenciando una brecha en métodos aplicables a disciplinas como el Derecho, donde los métodos numéricos tienen limitaciones. Para abordar esta necesidad, implementamos y validamos enfoques alternativos no cuantitativos, como el Chain of Thoughts y el Retrieval Augmented Generation. Estos métodos demostraron eficacia en la transparencia del razonamiento de la IA, facilitando el seguimiento del procesamiento de información y la generación de respuestas a través de formatos accesibles en software comercial. Puedes encontrar nuestro repositorio en el siguiente enlace: [https://github.com/pmalca/ulises\\_project](https://github.com/pmalca/ulises_project)

## I. INTRODUCCIÓN

“La Odisea” narra la travesía de Ulises en su regreso a Ítaca tras la guerra de Troya. Se trata de la historia de una auténtica travesía, donde el protagonista tiene que enfrentar la recurrente adversidad que le impide llegar a su destino. Es, esencialmente, la historia de un reto. Ulises finalmente consigue su objetivo tras un prolongado viaje que le exige emplear su mejor activo, su ingenio, para sortear los obstáculos que los dioses han puesto en su camino.

De manera similar, este trabajo propone un recorrido a través de los desafíos y oportunidades que ofrece la Inteligencia Artificial (en adelante, IA) aplicada al ámbito legal, centrando la atención en tres aspectos clave: la necesidad de explicabilidad en los sistemas de IA, el papel del Procesamiento de Lenguaje Natural<sup>1</sup> (en adelante, NLP) en el ámbito jurídico y un caso de uso concreto que permite analizar prácticas colusorias horizontales bajo la norma peruana de competencia, el Decreto Legislativo No. 1034, Ley de Represión de Conductas Anticompetitivas (en adelante, LRCA).

En la primera sección, abordaremos el concepto de *Explainable Artificial Intelligence* (en adelante, XAI). Asimismo, expondremos las principales discusiones que se han dado en este campo respecto a lo que constituye una ‘buena explicación’, y cómo es que determinar el usuario objetivo del sistema de IA es crucial para ello. En la segunda sección, se expondrá como el NLP, y su aplicación al Derecho (en adelante, Legal NLP), ofrece un marco teórico sólido para el desarrollo de estos sistemas.

En la tercera sección, se presenta al lector el concepto de *Legal Judgment Prediction* como una tarea de NLP aplicado a Derecho, que permite formular de forma coherente los objetivos de una IA. Además, se presentan los métodos más recientes desarrollados en este ámbito.

Por último, en la cuarta sección presentaremos un caso de uso que ejemplifica la implementación de estas tecnologías: un sistema diseñado para predecir el sentido de fallo en casos de prácticas colusorias horizontales, enmarcados en la LRCA. Este sistema no solo procesa información relevante, sino que también permite verificar y comprender cómo los datos analizados conducen a una predicción específica.

Para su elaboración, se ha utilizado un ‘Jupyter Notebook’, en el cual distintas funciones, algunas soportadas por un modelo de lenguaje a gran escala<sup>2</sup> (en adelante, LLM), realizan pequeños razonamientos jurídicos, utilizando como fuente de conocimiento externo las decisiones emitidas por las autoridades de competencia peruanas, y *zero-shot prompts* diseñados para fines específicos.

Este “viaje” hacia la integración de la IA en el ámbito legal es solo el comienzo. Nuestro enfoque inicial se limita a un área específica del Derecho de la Competencia, con resultados óptimos en español y circunscritos a la normativa peruana. Sin embargo, creemos firmemente que el desarrollo de este proyecto sienta las bases para la creación de sistemas más sofisticados que respeten criterios éticos, transparentes y orientados a las necesidades prácticas de los usuarios finales, aprovechando a su plenitud el vínculo entre la tecnología y el Derecho.

## I. ¿QUÉ ES LA INTELIGENCIA ARTIFICIAL EXPLICABLE (XAI)?

Para comprender la importancia de este proyecto y la metodología que hemos empleado para desarrollar este sistema, es fundamental comprender en qué consiste la XAI, campo en el cual se sitúa

---

<sup>1</sup> En el idioma inglés a este campo de investigación se le conoce como *natural language processing*.

<sup>2</sup> En el idioma inglés, a este tipo de modelos se les denomina *large language models*.

esta investigación. El objetivo de esta sección es proporcionar el itinerario mínimo para aproximarse a este, así como comprender sus objetivos.

XAI es un campo emergente de investigación que busca desarrollar métodos y técnicas para hacer que los sistemas de IA sean más comprensibles para sus distintos sujetos interesados (Maruthi et al, 2022, p. 3). Así, XAI tiene como objetivo reducir esta brecha al proporcionar explicaciones comprensibles que permitan a los usuarios entender los factores que influyen en las decisiones de los sistemas de IA.

La *explainability* o explicabilidad implica proporcionar información accesible y comprensible acerca de los factores y procesos de decisión que sustentan las predicciones, recomendaciones o decisiones que arroja un sistema de IA. Esto permite a los usuarios entender, cuestionar o impugnar los resultados, a la par que favorece a los desarrolladores y otros actores involucrados en la creación de IA, quienes se ven beneficiados al obtener un mayor entendimiento sobre el sistema (OECD, 2023, p. 32).

En ese sentido, la importancia de la XAI radica en: a) promover la confianza, facilitando la adopción de los resultados del modelo; b) garantizar la validez del contenido de estos resultados al identificar soluciones incorrectas o relaciones espurias; c) permitir una mayor aplicabilidad en el mundo real al ajustar las soluciones a problemas específicos; d) asegurar el cumplimiento de regulaciones al proporcionar trazabilidad en las decisiones; y e) facilitar la detección de sesgos no deseados (Zhou et al, 2024, p. 2).

Por otro lado, un concepto similar al de *explainability* es el de *interpretability*, que consiste en la capacidad de comprender los procesos internos de un modelo, revelando cómo toma sus decisiones mediante técnicas intrínsecas. Las técnicas intrínsecas se encuentran integradas en el modelo desde su creación y explican las propiedades internas del modelo de forma accesible para los humanos, permitiendo comprender cómo se toman las decisiones (Ali et al, 2023, p. 8). A efectos del presente trabajo, ambos conceptos serán tratados como sinónimos.

Sin embargo, el desarrollo de una explicación efectiva depende del nivel de conocimiento, experiencia y familiaridad del usuario con el sistema. La XAI emplea dos enfoques principales para resolver este problema (Zhou et al, 2024, p. 2): diseñar modelos intrínsecamente interpretables y proporcionar explicaciones externas que aclaren el proceso de toma de decisiones, destacando características relevantes, interacciones clave o posibles problemas.

Con relación a los usuarios de un sistema de IA y su relación con la explicabilidad, siguiendo a Munn & Pitman (2023, pp. 18-19), estos pueden ser divididos de la siguiente manera:

- **Practicantes** (Científicos de Datos e Ingenieros de *Machine Learning*, en adelante ML): Utilizan la explicabilidad durante la construcción y ajuste del modelo. Su objetivo es traducir las explicaciones en acciones concretas para mejorar el desempeño del modelo, como ajustar la arquitectura, los datos de entrenamiento o la estructura del *dataset*.
- **Observadores** (*Stakeholders* y Reguladores): Son personas ajenas a la construcción y despliegue del modelo. Incluyen *stakeholders* internos a la organización<sup>3</sup> y reguladores externos, interesados en entender y supervisar el modelo.
- **Usuarios Finales** (Expertos del Dominio y Usuarios Afectados):

---

<sup>3</sup> Un ejemplo de *stakeholder* serían los accionistas o miembros del directorio de una compañía.

- Expertos del dominio: Utilizan explicaciones como herramienta de apoyo en la toma de decisiones, aunque tengan poco conocimiento en ML.
- Usuarios afectados: Impactados por las predicciones del modelo, buscan explicaciones que permitan evaluar si la decisión fue justa y cómo modificar factores bajo su control para influir en futuras predicciones.

Por su parte, como ha sido previamente expuesto por Srihari (2024, pp. 8-9), los métodos XAI pueden ser clasificados en dos categorías, de acuerdo al momento en el que se genera la explicación:

- 1 *Post-hoc*: Las explicaciones son generadas a partir de sistemas de IA existentes, las cuales pueden obtenerse perturbando variables de entrada en mapas de calor o mediante mecanismos de atención en modelos de aprendizaje profundo para contextualizar las decisiones. Ejemplos incluyen Sensitivity Analysis (en adelante, SA), Layer-wise Relevance Propagation (en adelante, LRP) y LIME. Se utilizan técnicas como perturbación de variables y mecanismos de atención en modelos profundos.
- 2 *Ante-hoc*: Los sistemas son diseñados desde el principio para generar explicaciones. Ejemplos son RETAIN y *Bayesian Deep Learning* (en adelante, BDL). Utilizan métodos como la reducción de dimensiones de salidas de capas o enfoques neuro-simbólicos.

Con relación a estos métodos, es necesario destacar que su resultado y explicaciones se diseñan y modulan considerando como estándar el conocimiento de los expertos en IA (siendo explicables “para ellos”). Este estándar descuidaría las necesidades de una comunidad más amplia, como lo son los observadores y los usuarios finales.

Además, se ha cuestionado la eficacia de las explicaciones actuales para los usuarios finales expertos del dominio, es decir, aquellos que son expertos en un campo específico, pero carecen de conocimientos técnicos en IA (Bhattacharya, 2024, p. 2). Dicha aproximación refleja que el estado actual de XAI está centrada más en los aspectos técnicos y de diseño desde la perspectiva de los practicantes, que en los desafíos prácticos de su implementación (Vainio-Pekka et al, 2023, p.26). De este modo, el enfoque predominante en los estudios de XAI se dirige a la propia arquitectura del modelo, orientándose principalmente a “usuarios técnicos” (tales como ingenieros en ML o *Deep Learning*), o a quienes poseen un conocimiento profundo de la matemática detrás de los modelos. Este enfoque deja de lado a otros sujetos interesados en la explicabilidad del sistema de IA, pero que usualmente no son especialistas en la materia, como aquellos formados en Derecho.

Asimismo, pocos estudios se enfocan en evaluar si las explicaciones generadas son comprensibles y aceptables para los usuarios. Esta evaluación es crucial para garantizar que XAI promueva la responsabilidad, pues solo el escrutinio abierto permite explorar las auténticas limitaciones de un modelo. Aún existe una brecha en la investigación sobre cómo los humanos interpretan los datos, lo que dificulta la creación de sistemas XAI que coincidan con su comprensión (Baker & Xiang, 2023, p. 27).

Aunado a lo anterior, y a pesar de su alto rendimiento, los modelos de lenguaje son criticados por su falta de interpretabilidad, atribuida a su arquitectura compleja y a la opacidad de los modelos de *Deep Learning*<sup>4</sup>. Esta falta de claridad genera dudas sobre cómo toman decisiones y sobre la confianza en ellas, especialmente en aplicaciones críticas donde la transparencia es crucial (Maruthi et al, 2022, p. 3).

---

<sup>4</sup> Este problema referido a la opacidad también es conocido como ‘black-box’. Se trata de un modelo ‘black-box’ cuando no existen métodos adecuados para explicar su funcionamiento interno. En *Deep Learning* el problema surge cuando, por ejemplo, no es factible identificar las neuronas que son activadas dentro de su arquitectura.

Por otro lado, la IA Generativa (en adelante, IAG) ha ganado relevancia en los últimos años, pero enfrenta problemas como sesgos y desinformación. Un enfoque XAI puede ayudar a que los sistemas de IAG sean más responsables, proporcionando explicaciones sobre cómo el modelo entendió una pregunta y generó una respuesta (Baker & Xiang, 2023, p. 24). Esto permite someter los resultados de la IAG al escrutinio de un elenco de actores más amplio, al viabilizar la trazabilidad de sus resultados y facilitar la identificación de eventuales situaciones problemáticas en el proceso de generación de respuestas.

Debido a la novedad de estas tecnologías, no existe un consenso respecto a los métodos que deben ser adoptados para la IAG. No obstante, Meller propone como herramientas específicas para mejorar la explicabilidad a las técnicas denominadas *Chain of Thoughts* (en adelante, CoT) y *Retrieval Augmented Generation* (en adelante, RAG). Aunque ambas están diseñadas principalmente para mejorar la calidad de las respuestas de los modelos, su ventaja adicional radica en aumentar su explicabilidad y confianza (Meller, 2024, p. 47).

En este contexto, y como se desarrollará a mayor profundidad en las siguientes secciones, nuestro sistema de IA estará diseñado para generar explicaciones dirigidas a un usuario final experto del dominio: un profesional especializado en Derecho de la Libre Competencia en Perú. La elección de una metodología *ante-hoc* responde a la necesidad de que el sistema proporcione explicaciones intermedias de forma autónoma, detallando la información procesada que derivó en las respuestas entregadas por el sistema.

Nuestro sistema integra IAG en algunos de sus algoritmos, destacando el uso de LLMs. En ese orden de ideas, técnicas como CoT y RAG cumplen una función crucial al actuar como opciones no matemáticas para desarrollar sistemas explicables. Nótese que estas no solo mejoran la calidad y relevancia de las respuestas generadas, sino también, permiten ofrecer explicaciones claras y accesibles.

## II. UNA APROXIMACIÓN AL LEGAL NLP

Lo que guardan en común el jurista, el abogado, el juez y el funcionario público, sea cual sea la rama del Derecho a la que se dediquen, es el uso preponderante del texto en lenguaje natural para su labor. En ese sentido, los algoritmos de Legal NLP, al ofrecer un método computacional para analizarlo, son útiles para la investigación jurídica, ya que amplifican la variedad de metodologías, específicamente, para estudios empíricos del Derecho (Katz et al, 2023).

Hay muchas aplicaciones basadas en Legal NLP, como los *document embeddings* para decisiones judiciales que facilita la posibilidad de generar hipótesis cuantitativas y empíricamente comprobables sobre el Derecho (Ash & Chen, 2018), la extracción de derechos y deberes de contratos laborales (Ash et al, 2020), *word embeddings* para una rama del derecho en específico (Chalkidis & Kampas, 2019), el resumen de decisiones de las cortes de los EE. UU., o legal judgment prediction, comparación de casos similares y sistemas de preguntas y respuestas (Zhon et al, 2020).

Los corpus, bases de datos con información de relevancia jurídica, son cruciales para implementar los algoritmos de Legal NLP. Cabe señalar que, como ocurre con otros campos especializados, el texto jurídico tiene características diferentes a un corpus de texto de propósito general, como lo sería el vocabulario de un lenguaje o la semántica. Esto crea un desafío adicional, ya que las nuevas tecnologías de IA, como los *large pretrained transformers*, necesitan grandes conjuntos de datos especializados para ser útiles en el ámbito legal, y estos requieren miles de anotaciones realizadas por expertos (Bambroo & Awasthi, 2021).

Hay muchas contribuciones para llenar este vacío. El Cambridge Law Corpus pretende ser una base de datos de alta calidad de casos judiciales del Reino Unido, Inglaterra y Gales (Östling et al, 2023). Existen otras bases

de datos diseñadas para fines más específicos. LEDGAR, por ejemplo, pretende ser un corpus diseñado para la clasificación textual de disposiciones legales en los contratos (Tuggener et al, 2020). Otros corpus, como aquel dedicado a fallos de los tribunales franceses, están diseñados para la clasificación de textos (Sulea et al, 2017), el conjunto de datos VICTOR, para la clasificación de tipos de documentos y asignación de temas (Luz de Araujo et al, 2020), o el conjunto de datos CaseHOLD, que está diseñado para identificar *overrulings* y clasificar términos de servicios en decisiones estadounidenses (Zheng et al, 2021).

Cabe destacar que la comunidad de Legal NLP participa activamente en el desarrollo de *benchmarks* para evaluar el desempeño de los algoritmos en tareas que involucran razonamiento legal. Si bien los *benchmarks* no son inmunes a las críticas (Raji et al, 2021), este esfuerzo continuo debe ser aprovechado como parámetro para evaluar la eficacia de los algoritmos en tareas que requieren capacidades de razonamiento legal.

Sobre el particular, hemos identificado dos *benchmarks* relevantes para nuestra investigación. En primer lugar, LexGLUE (Chalkidis et al, 2021), que es una colección de bases de datos diseñados para evaluar sistemáticamente el rendimiento del modelo en diversas *legal natural language understanding tasks*.

El segundo es LegalBench (Guha et al, 2023), que se centra específicamente en evaluar las capacidades de los LLM para realizar los cuatro pasos del método de análisis legal "IRAC"<sup>5</sup>:

- Identificación del problema legal (*issue-spotting*): Analizar los hechos para detectar cuestiones legales, ya sea una pregunta específica o un área del derecho relevante. En algunos casos, se les indica explícitamente el problema; en otros, deben inferirlo.
- Identificación de las normas legales aplicables (*rule-recall*): Determinar las reglas que establecen las condiciones necesarias para que se arribe a un resultado específico. En EE.UU., estas normas provienen de diversas fuentes, como la Constitución, leyes federales y estatales, regulaciones y jurisprudencia.
- Aplicación de las normas a los hechos (*rule-application*): Analizar los hechos relevantes en relación con las normas, determinando cómo influyen en el sentido del fallo. Esto puede incluir citar precedentes para comparar similitudes o diferencias que orienten la resolución del caso.
- Conclusión legal (*rule-conclusion*): Basándose en la aplicación de la norma a los hechos, determinar el sentido del fallo.

Con relación al uso del IRAC en la profesión legal y siguiendo a Dobson (2020, p.1) y a Sherkow (2024, p. 39), este un sistema de análisis, generalmente empleado para la redacción de respuestas en exámenes de derecho y/o documentos legales (consultas, memorándums, entre otros)<sup>6</sup>. La Columbia Law School refiere que la redacción legal tiene una estructura organizacional muy específica, que puede ser descrita como IRAC o una de sus variaciones (s/f, p. 1). En la misma línea, Bench-Capon (2020, p. 74) refiere que el método IRAC, es una metodología para el análisis legal ampliamente enseñada en las facultades de derecho de Estados Unidos como una forma de responder preguntas hipotéticas presentadas durante la enseñanza a través del método socrático y en exámenes.

---

5 El método IRAC es una estructura utilizada en el análisis legal que consiste en identificar el problema (*Issue*), aplicar la norma (*Rule*), analizar cómo se aplica al caso (*Application*) y llegar a una conclusión (*Conclusion*). Este enfoque ayuda a organizar el razonamiento jurídico de manera clara y coherente.

6 Parece no haber certeza respecto a quién "creó" el método IRAC, siendo que algunas personas lo atribuyen al Ejército de Estados Unidos de América para facilitar el entrenamiento de reclutas levados para combatir en la Segunda Guerra Mundial.

En conclusión, el Legal NLP constituye una base sólida para el desarrollo de soluciones computacionales que aborden problemas que involucran razonamiento jurídico. La capacidad de estos algoritmos para analizar textos legales de manera eficiente permite explorar nuevas metodologías, especialmente en el ámbito de los estudios empíricos del Derecho, y fomenta aplicaciones prácticas que van desde la clasificación de documentos legales hasta la predicción de fallos judiciales.

Además, el método IRAC, como se verá más adelante, se posiciona como una herramienta clave para estructurar soluciones computacionales en el ámbito jurídico. Al descomponer problemas legales en pasos claros y definidos, el IRAC facilita la implementación de sistemas que son tanto efectivos como comprensibles. La combinación del Legal NLP y el IRAC no solo amplía las posibilidades de investigación y desarrollo en el campo legal, sino que también establece una metodología sistemática para enfrentar desafíos complejos con soluciones computacionales accesibles y precisas.

### III. ¿CÓMO PREDECIR LOS SENTIDOS DE FALLOS?: EL *LEGAL JUDGMENT PREDICTION* COMO TAREA LEGAL NLP

El *Legal Judgment Prediction* (en adelante, LJP) es la tarea Legal NLP que utiliza como insumo la descripción de los hechos del caso, con el objetivo de predecir el resultado de su juzgamiento. Los modelos basados en esta tarea pueden asistir a los abogados y ciudadanos, al reducir los costos legales y promover el acceso a la justicia. Los abogados y jueces pueden utilizarlos para estimar la probabilidad de ganar un caso y arribar a valoraciones más consistentes e informadas (Chalkidis, Androutsopoulos & Aletras p. 1).

Los primeros métodos LJP dependían de la extracción manual de características, un proceso simple pero costoso. Con el auge del ML, los investigadores comenzaron a abordar el problema de manera más eficiente mediante métodos basados en datos, que requieren menos intervención humana (Wu et al, 2023, pp. 12061 – 12062).

Asimismo, es importante reconocer la relevancia de la explicabilidad e interpretabilidad de los resultados en el LJP. Existen diversas aproximaciones a la explicabilidad en la literatura de LJP, las cuales a su vez dependen de la manera en que se resuelva la tarea de predecir el resultado del juicio. Por ejemplo, hay sistemas de aprendizaje automático que tratan la LJP como una tarea de clasificación y generan explicaciones sobre las características importantes<sup>7</sup>. Otros sistemas abordan la LJP como una tarea de generación de texto, y abordan la explicabilidad en base a un conjunto de pasos de razonamiento o silogismos legales con conclusiones (Medvedeva & McBride, 2023, p. 79).

Además, las aproximaciones interpretables se dividen en *pre-explicación*, donde primero se generan explicaciones que guían las predicciones, y *post-explicación*, donde las predicciones se obtienen primero y luego se generan explicaciones basadas en la relevancia de las sentencias o la reconstrucción de la lógica judicial del fallo (Feng, Li & Ng, 2022, p. 5464).

No obstante, aunque el enfoque en la explicabilidad es positivo, la mayoría de las explicaciones no cumplen con tres criterios esenciales: estar fundamentadas en el Derecho, relacionarse con predicciones de futuros fallos y ser evaluadas por métodos adecuados. Esto representa una limitación significativa para la evaluación, utilidad real y los impactos de los sistemas LJP (Medvedeva & McBride, 2023, p. 79).

---

<sup>7</sup> En *machine learning* se utiliza el término *feature* o característica para hacer referencia a una parte del *input*. Por ejemplo, una característica importante en las tareas NLP serían las palabras o *token* que más influyen en el *output* generado.

Las aproximaciones actuales de interpretabilidad en LJP no han tenido éxito, ya que los resultados empíricos muestran que la generación de explicaciones, ya sea en un enfoque de pre- o *post-explicación*, afecta negativamente a los resultados de LJP. Esto se debe a que LJP y la generación de explicaciones suelen ser aprendidos conjuntamente, lo que aumenta la complejidad del aprendizaje y permite que la generación de explicaciones influya en los resultados de LJP<sup>8</sup>, incluso en los enfoques de post-explicación. Además, otro problema es que estas aproximaciones no garantizan que las explicaciones y los resultados de las predicciones sean consistentes (Feng, Li & Ng, 2022, p. 5465).

Considerando ello, a continuación, explicaremos algunos de los métodos y aspectos a considerar en el funcionamiento de los sistemas LJP:

## i. CoT y LJP

El *prompt* CoT, propuesto por Wei et al (2023), permite descomponer problemas complejos en pasos intermedios, asignando más capacidad de cómputo al razonamiento. Además, facilita interpretar el comportamiento del modelo y detectar errores en su razonamiento, no obstante que caracterizar completamente estos cálculos sigue siendo un desafío. Esta técnica es útil en tareas como problemas matemáticos, razonamiento común y manipulación simbólica, con potencial para cualquier tarea resoluble mediante lenguaje. Finalmente, puede inducirse fácilmente en modelos grandes mediante ejemplos en *few-shot prompting*<sup>9</sup> (Wei et al, 2023, p. 3).

En el ámbito del Legal NLP, tenemos una adaptación de este enfoque denominado *Legal Syllogism Prompting*. Propuesto por Jiang y Yang (2023), este método reformula el LJP al pasar de una clasificación a una generación de texto, estructurando la entrada según las premisas del silogismo: la premisa mayor como el texto de la ley, la premisa menor como los hechos del caso y la conclusión como el juicio resultante. Este enfoque no solo activa la capacidad de razonamiento deductivo de los LLM, sino que también genera salidas en formatos estructurados que facilitan la verificación posterior.

## ii. RAG y LJP

Un segundo aspecto que un sistema LJP debe tener en consideración es el uso de casos previos para brindar la respuesta. En el sistema de Common Law, el precedente es la base para resolver una controversia, así como también es una referencia clave en el sistema del Civil Law<sup>10</sup>. Por ello, la recuperación de precedentes es una tarea importante en la IA destinada a la actividad legal.

Con relación a su aplicación para el LJP, existen dos modelos principales para esta tarea: aquellos basados en conocimiento experto, que representan casos legales mediante subelementos diseñados, y los basados en NLP, que mayormente convierten el texto en *embeddings* para calcular similitudes (Wu et al, 2023, p. 12062).

---

8 En este tipo de tareas el modelo genera como *output* no solo la predicción del caso, sino también, los motivos que sustentan esa predicción. Por ejemplo, si el modelo tiene que evaluar si los hechos constituyen una infracción a la Norma X, primero debe producir los criterios establecidos en la citada norma, y luego, explicar cómo los aplica al caso, antes de responder si constituye o no una infracción. Así, el problema que debe ser resuelto no solo implica encontrar los patrones entre la descripción de los hechos del caso y la predicción (i.e. infracción o no infracción), sino también, la relación que existe entre estos elementos y el conjunto de palabras (o *tokens*) que componen la explicación.

Cabe destacar que el estudio fue elaborado de forma previa al lanzamiento de ChatGPT, cuando los modelos de lenguaje de esa época adolecían de varias limitaciones en cuanto a la generación de texto.

9 El *few-shot prompting* es una técnica que consiste en incluir ejemplos de la tarea dentro del *prompt*. Este enfoque fue introducido en el artículo Language Models are Few-Shot Learners, cuyo enlace es el siguiente: <https://arxiv.org/pdf/2005.14165>

10 Para una mejor comprensión sobre la diferencia entre ambos sistemas, así como su relación con el Derecho de la Competencia, puede consultarse el siguiente artículo: [http://centrocompetencia.com/wp-content/uploads/2024/06/PDF-VERSION-2-Ignacio-Peralta-junio-2024-v-2-002\\_1.pdf](http://centrocompetencia.com/wp-content/uploads/2024/06/PDF-VERSION-2-Ignacio-Peralta-junio-2024-v-2-002_1.pdf)

Por otro lado, Wu et al definieron como problema de LJP el siguiente: 'Dada una descripción  $f$ , nuestra tarea es obtener y comprender los precedentes  $P$ , después, predecir el resultado  $j=(a,c,t)$ '<sup>11</sup> (2023, p. 12062). Es por tal razón que proponen un marco denominado *Precedent-Enhanced Legal Judgment Prediction*, que combina la capacidad de los LLM y los modelos especializados para mejorar la recuperación y comprensión de precedentes, demostrando su efectividad en los experimentos conducidos (2023, p. 12608).

Peng y Cheng presentaron el *framework Athena*. Este utiliza RAG para mejorar el rendimiento de los LLM en la predicción. Al integrar una base de conocimiento externa mediante recuperación semántica<sup>12</sup>, Athena aborda limitaciones clave de los LLM, como las alucinaciones y la falta de conocimiento específico, estableciendo un camino hacia decisiones legales asistidas por IA más confiables. Así, recomiendan que futuras investigaciones se centren en optimizar el subsistema de recuperación y refinar la lógica de inferencia de los LLM, incluyendo flujos de trabajo complejos, para potenciar aún más las capacidades de Athena en la predicción de decisiones legales (2024, p.13).

Asimismo, Yue et al desarrollaron DISC-LawLLM, un sistema legal inteligente diseñado para ofrecer diversos servicios legales. Para mejorar la fiabilidad de las respuestas, se integró un módulo de recuperación de información externo (2023, p.9).

También, los autores destacan como principales desafíos para desarrollar el sistema: (1) Alta exigencia de capacidad de razonamiento en temas legales, y (2) Necesidad de capacidad de recuperación e inferencia de conocimiento legal externo (Yue et al, 2023, p.2).

Otro estudio que utiliza casos previos para mejorar la precisión del LJP es el conducido por Han & Dou, en el que proponen *el Case Retrieval framework* para LJP. Este *framework* integra un módulo de recuperación de casos históricos para simular el proceso judicial real de consulta de precedentes antes de emitir un fallo. Así, al partir de la premisa de que los jueces suelen consultar casos históricos antes de emitir un fallo, se diseñó un modelo de recuperación de casos históricos para simular este escenario.

Para un caso actual, se recuperan los *top-k* casos históricos<sup>13</sup> más similares y se obtienen sus representaciones vectoriales. Luego, se concatena el vector promedio de estos casos con el vector de la descripción de los hechos para predecir los resultados del juicio. Los resultados experimentales demostrarían la eficacia del método (Han & Dou, 2023, p. 809).

## IV. PROYECTO ULISES: IMPLEMENTACIÓN DE UN SISTEMA PREDICTOR DE CASOS EN MATERIA DE PRÁCTICAS COLUSORIAS HORIZONTALES PARA EL DERECHO DE LA LIBRE COMPETENCIA PERUANO

Habiendo definido los aspectos clave del campo en el cual se desarrolla el proyecto y hecho el estudio de la literatura especializada en la materia, procederemos a describir los resultados del proyecto y su rendimiento<sup>14</sup>.

---

11 La notación es la siguiente. ' $a$ ' hace referencia a la norma; ' $c$ ', a los cargos; y ' $t$ ' al tiempo estimado de prisión.

12 Con recuperación semántica se hace referencia a aquella que se ejecuta mediante *embeddings*.

13 La elección del "top-k" óptimo no es universal. Este dependerá de la tarea en cuestión, lo cual incidirá en el enfoque computacional que sea el más adecuado para recuperar los casos similares. Cabe destacar que el método elegido también puede incorporar conocimiento de un abogado experto.

14 Es importante precisar que Ulises es un sistema creado con fines estrictamente académicos a propósito de los temas descritos en este artículo. Su uso no corresponde ni sustituye la asesoría legal especializada y los resultados que se arrojen se integran a los documentos de respaldo que se generan, los cuales pueden ser revisados para validar y examinar el proceso de funcionamiento del sistema y las respuestas a las que llega.

Para esto, partiremos de una revisión general del funcionamiento del sistema peruano de defensa de la libre competencia (en el cual está basado el sistema de Ulises), a fin de que el lector pueda contar con un contexto adecuado que le permita interpretar las respuestas que arroje el sistema, para luego pasar a la descripción propia del sistema y la metodología empleada para su funcionamiento.

### a. El régimen peruano de defensa de la libre competencia

Perú es una jurisdicción con un sistema de defensa de la competencia relativamente joven, con poco más de tres décadas y herramientas de *enforcement* que funcionan bajo un sistema administrativo sancionador. Para efectos de esta sección, nos centraremos en la institucionalidad peruana de libre competencia considerando su composición actual y las disposiciones relevantes de la LRCA en materia de control de conductas anticompetitivas.

Bajo el ordenamiento peruano coexisten dos agencias peruanas que tienen la habilitación legal para aplicar las disposiciones de la LRCA: el Instituto Nacional de Defensa de la Competencia y de la Propiedad Intelectual (en adelante, INDECOPI) y el Organismo Supervisor de la Inversión Privada en Telecomunicaciones (en adelante, OSIPTEL). Ambas entidades tienen la facultad legal de emplear la norma para hacer *enforcement* de libre competencia, contando con órganos especializados para investigar, perseguir y sancionar a los agentes económicos que ejecuten infracciones contra la libre competencia.

La diferencia sustancial entre ambas agencias y sus facultades está en su ámbito de competencia por materia. Mientras que el INDECOPI es la agencia de competencia “general” del Perú, con la habilitación para conocer los casos relacionados a todos los sectores de la economía; OSIPTEL tiene su competencia acotada únicamente para aquellos casos relacionados con la prestación del servicio público de telecomunicaciones, siendo este el único sector donde opera esta excepción (Falla, 2019). Así, si bien ambas agencias cuentan con estructuras similares, la competencia de cada una se delimita en función al mercado afectado por la conducta investigada, siendo que ambas agencias coordinan sus actividades para evitar divergencia entre sus criterios y decisiones (OCDE y BID, 2018, p. 108).

A nivel de INDECOPI, el órgano instructor encargado de investigar e impulsar los procedimientos es la Dirección Nacional de Investigación y Promoción de la Libre Competencia (antes Secretaría Técnica de la Comisión de Defensa de la Libre Competencia). Por su parte, el órgano resolutorio, encargado de *decidir* los casos es la Comisión de Defensa de la Libre Competencia en primera instancia y la Sala Especializada en Defensa de la Competencia del Tribunal del INDECOPI, en segunda instancia. De otro lado, el OSIPTEL tiene como órgano instructor a la Secretaría Técnica de los Cuerpos Colegiados, como órgano resolutorio de primera instancia a los Cuerpos Colegiados Permanentes y como segunda instancia al Tribunal de Resolución de Controversias.

Esta ha sido la descripción de los organismos competentes en sede administrativa. Bajo el régimen peruano, es posible cuestionar las decisiones de la autoridad de competencia ante el Poder Judicial, bajo una acción contencioso-administrativa que le permite al juez evaluar la legalidad de la decisión emitida por el órgano e incluso resolver sobre el fondo de plena jurisdicción.

Debe considerarse que la LRCA es una norma aprobada el año 2008, recogiendo muchas de las reglas desarrolladas por el INDECOPI a lo largo de su historia. La referida norma privilegia los instrumentos públicos para su *enforcement*, procurando su cumplimiento principalmente a través de sanciones administrativas que se imponen respecto de conductas anticompetitivas típicamente sancionadas en la mayoría de jurisdicciones, como el abuso de posición de dominio o las prácticas colusorias (Bullard & Valencia-Dongo, 2022, pp. 61-64).

Considerando que el régimen peruano de control de conductas opera bajo las reglas del procedimiento administrativo sancionador, la posibilidad de la autoridad para investigar, analizar y sancionar conductas anticompetitivas depende de su previa y necesaria tipificación como infracciones en la ley.

El primer grupo de conductas sancionables bajo la LRCA son las prácticas de abuso de posición de dominio, que consisten en la conducta desplegada por un agente económico que cuenta con posición dominante en un mercado relevante determinado.

Bajo el régimen peruano, sólo existirá abuso de posición de dominio cuando una empresa en tal posición ejercite indebidamente su poder de mercado con el fin de perjudicar a sus competidores para poder obtener beneficios y causar perjuicios a los demás agentes económicos. La norma peruana solo sanciona las conductas de abuso de posición de dominio con efectos exclusorios, esto es, que se encuentren destinadas a afectar el proceso competitivo en sí al obstaculizar la permanencia de otros agentes en el mercado o excluirlos directamente. De este modo, queda completamente excluida la posibilidad de sancionar abusos explotativos (Patrón, 2008, p. 133)

Algunas de las conductas tipificadas expresamente como prácticas anticompetitivas de este tipo son la negativa injustificada de venta, la aplicación de condiciones comerciales desiguales a situaciones equivalentes, los empaquetamientos, la obstaculización injustificada a la entrada o permanencia de un competidor en una asociación, el uso abusivo de procesos legales, la suscripción de contratos de exclusividad o cláusulas de no competencia que no sean justificadas o la negativa a proveer bienes o prestar servicios, entre otros.

El segundo grupo de conductas son las prácticas colusorias, que a su vez se subdividen en prácticas colusorias horizontales y verticales. Las prácticas colusorias horizontales consisten en la coordinación entre dos o más agentes competidores que tiene por objeto o efecto restringir, suprimir o controlar la competencia real o potencial que existe en un mismo nivel de actividad económica (Borda, 2020, p. 190). Por su parte las prácticas colusorias verticales consisten en acuerdos que apuntan al mismo objetivo, pero que son realizados por agentes que operan en distintos niveles de la cadena de producción.

Entre las prácticas tipificadas como prácticas colusorias se encuentran la fijación concertada de forma directa o indirecta de precios u otras condiciones comerciales, la limitación de la producción, ventas, el reparto de clientes o proveedores, la subordinación de la celebración de contratos a la aceptación de prestaciones adicionales, la aplicación de condiciones desiguales a prestaciones equivalentes, la concertación injustificada de venta exclusiva, el establecimiento de posturas y abstenciones en licitaciones, entre otras.

Además, la LRCA cuenta con dos reglas de análisis aplicables a las conductas anticompetitivas. Siguiendo a Quintana, podemos definir la regla de prohibición absoluta como una prohibición automática, que permite a la autoridad sancionar una práctica por el solo hecho de existir y con independencia de un análisis de los perjuicios que ha ocasionado; por otro lado tenemos la regla de prohibición relativa, que consiste en un criterio de prohibición que se adopta en función de los efectos de la conducta, de tal forma que su ilegalidad dependerá de la existencia de efectos dañinos en el mercado, ponderados con los beneficios ocasionados (2013, p. 38).

Solo cuatro conductas pertenecientes al grupo de prácticas colusorias horizontales se encuentran sometidas a la prohibición absoluta: (i) La fijación de precios u otras condiciones comerciales o de servicio; (ii) La limitación de la producción o las ventas, en particular por medio de cuotas; (iii) El reparto de clientes, proveedores y zonas geográficas; y (iv) El establecimiento de posturas o abstenciones en las licitaciones, concursos o cualquier otra forma de contratación pública, cuando además dichas prácticas sean *intermarca* (esto es, realizadas respecto a bienes o servicios producidos o distribuidos por agentes económicos distintos o “de distinta marca”) y no sean accesorias a ningún acuerdo lícito. El resto de prácticas colusorias, y todas las

modalidades de abuso de posición de dominio, se encuentran sometidas a la regla de la razón, de manera que para su análisis la autoridad deberá ponderar las eficiencias generadas por la conducta respecto de los eventuales perjuicios de esta.

Las decisiones del INDECOPI y el OSIPTEL se materializan en decisiones administrativas plasmadas en resoluciones. Estas pueden disponer la absolució del imputado o la declaraci3n de su responsabilidad, lo que a su vez puede implicar la subsecuente imposici3n de sanciones, (como multas o amonestaciones) y medidas correctivas, que son mandatos de cumplimiento obligatorio para los agentes econ3micos.

Con esto claro, pasaremos a exponer c3mo es que las decisiones que producen las instituciones descritas son abordadas por nuestro sistema (Ulises). Sobre el particular, este se encuentra enfocado en predecir si una serie de hechos constituyen una pr3ctica colusoria horizontal sancionable bajo la legislaci3n peruana.

## b. Directivas para implementar un sistema predictor de casos (LJP) para el Derecho Peruano de la Libre Competencia

Nuestro sistema toma en consideraci3n las siguientes recomendaciones propuestas por Medvedeva & McBride (2023, p. 84), para la elaboraci3n de sistemas LJP:

Ítem	Objetivo	Aplicaci3n
Usuario final	Definir al usuario final del sistema.	Profesionales especialistas en Derecho y Polític as de Competencia.
Aplicaci3n del sistema	Definir la aplicaci3n y prop3sito del sistema dentro del proceso legal.	El prop3sito del sistema es brindar una primera aproximaci3n al car3cter anticompetitivo, o no, de las pr3cticas comerciales identificadas por el usuario final.
	Determinar la etapa del proceso en el que el sistema podría ser utilizado.	El sistema puede ser utilizado en cualquier etapa del proceso.
Evaluaci3n de la data	Evaluar si la data de prueba (test) refleja los datos disponibles para el usuario en la etapa relevante del procedimiento legal de acuerdo con la aplicaci3n definida.	El sistema es testeado con ejemplos de conductas anticompetitivas analizadas por la autoridad de competencia y otras generadas a trav3s de data sint3tica. La extensi3n y descripci3n de los hechos pueden ser similares a las consultas del usuario final.
Explicabilidad	Evaluar si el sistema necesita proporcionar explicaciones para sus predicciones dada la aplicaci3n y, de ser así, especifique la naturaleza de las explicaciones requeridas.	Sí. El sistema necesita brindar explicaciones respecto a la aplicaci3n de la ley de competencia a los hechos del caso.
Consideraciones éticas	Identificar posibles beneficios o daños relacionados con el rendimiento del sistema, como predicciones (incorrectas) que afectan los procedimientos legales, la vida de las personas y las protecciones legales.	Este sistema no está pensado para automatizar la toma de decisiones, sino como una ayuda para el tomador de decisiones.
	Considerar las posibles repercusiones más amplias de las opciones de diseño del sistema, incluidos los impactos en las personas, las leyes, los procedimientos legales y la sociedad en general.	Los códigos, así como las bases de datos, son públicos, por lo que cualquier sesgo u otra falla puede ser identificada por la comunidad de profesionales especialistas en Derecho y Polític as de Competencia.

Asimismo, los códigos y bases de datos necesarias para correr este experimento se pueden encontrar en el siguiente repositorio en GitHub<sup>15</sup>:

## i. El usuario del sistema

LJP es una tarea de clasificación de texto en el ámbito del NLP, que busca predecir el resultado de decisiones judiciales a partir de descripciones textuales de los hechos del caso. En nuestro enfoque, los datos de entrada (*input*) consisten en estas descripciones, cuya extensión y contenido pueden variar considerablemente según el perfil del redactor. Por ejemplo, un usuario sin formación jurídica podría proporcionar descripciones vagas o incompletas; un profesional del derecho no especializado en libre competencia podría omitir aspectos técnicos esenciales; mientras que un especialista en libre competencia probablemente presentará una narrativa más estructurada y con mayor precisión técnica.

El sistema está diseñado principalmente para ser utilizado por especialistas en derecho de libre competencia, ya que los casos evaluados requieren un conocimiento previo profundo del área y una capacidad para identificar y valorar medios probatorios relevantes. Esto asegura que las descripciones de los hechos sean completas y adecuadas para probar el sistema, permitiendo una evaluación más precisa y confiable de su desempeño en escenarios reales.

## ii. Generación de un *benchmark* o *dataset* de evaluación para el Derecho de la Libre Competencia peruano

El Derecho de la Libre Competencia en el Perú carece de *datasets* de evaluación o *benchmarks* específicos, lo que nos llevó a desarrollar uno propio. Es importante señalar que nuestro sistema se inspira en la versión tradicional de LJP, por lo que la evaluación se realizará en función de la capacidad de la máquina para predecir correctamente el valor de los datos de salida (*output*).

Data de entrada	Data de salida
Descripción de los hechos del caso	Infracción / No infracción

Para elaborar nuestro *benchmark*, hemos decidido incluir tanto información textual relacionada con la práctica real como datos sintéticos. En cuanto a la primera, en el documento “Guía de Asociaciones Gremiales” (2021), el INDECOPI presenta ejemplos de hechos o acciones que constituirían una violación a las normas de competencia, junto con una contextualización de las razones por las cuales tales prácticas están prohibidas.

En cuanto a la segunda, hemos seguido las recomendaciones de Guo & Chen (2024) para la generación de datos sintéticos. Específicamente, para generar estos datos, se utiliza como entrada al modelo de lenguaje la información de salida que el sistema debe predecir, a fin de generar la data de entrada que será utilizada por el sistema:

---

15 El enlace al repositorio de GitHub es el siguiente: [https://github.com/pmalca/ulises\\_project](https://github.com/pmalca/ulises_project)

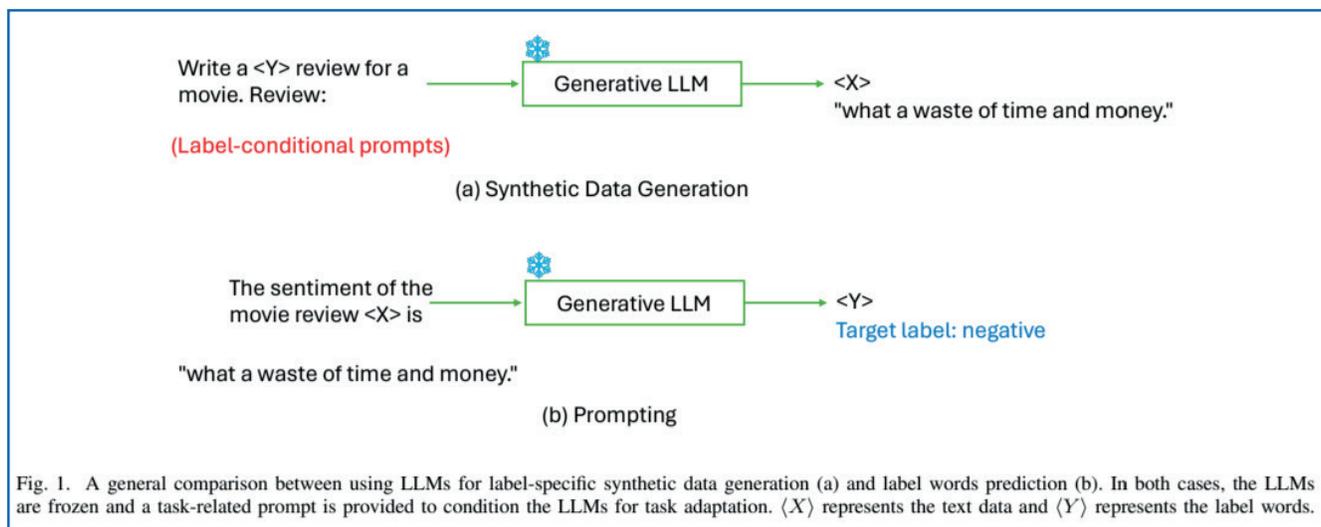


Gráfico 1: Método correcto para generar data sintética con LLM

Extraído de *Generative AI for synthetic data generation: Methods, challenges and the future*, p. 2

Así, nuestro *dataset* de evaluación está compuesto por once descripciones, cada una acompañada de una columna con valores booleanos que indican si hubo o no una infracción a la ley de competencia. Además, incluye otra columna con las razones que sustentan la existencia o ausencia de dicha infracción. Este se encuentra almacenado en el archivo 'dataset\_test.xlsx'.

### c. IRAC como referencia para diseñar nuestro sistema de LJP

#### i. Introducción al IRAC y su uso en Legal NLP

De acuerdo con Bench-Capon, la aplicación del método IRAC responde a la necesidad de intentar producir explicaciones más naturales de casos legales, y se enfoca en cómo se les enseña a los abogados a discutir casos (2020, p. 74). Además, precisa que el punto clave sobre IRAC es que está específicamente adaptado al caso en cuestión: indica lo que es importante y diferente sobre el caso particular en discusión, y utiliza los hechos específicos del caso para aplicar la regla general. Esto difiere de las formas estándar de explicación de los resultados de casos encontrados en IA y Derecho, que abordan cada posible problema en el dominio, oscureciendo el punto clave, y terminan en factores generalmente aplicables en lugar de hechos específicos (2020, p. 75).

El método IRAC ha sido elegido para el análisis de casos legales a través de LLM. Por ejemplo, Kang et al (2023, p. 1) emplearon ChatGPT para realizar un análisis IRAC de un conjunto de escenarios relacionados a la Ley de Contratos de Malasia y la Ley Social Australiana para Niños Dependientes, arrojando un análisis IRAC individual para cada escenario, en tanto consideran que el IRAC es la metodología de análisis más popular empleada por profesionales legales y facultades de derecho.

Asimismo, como señalamos en la sección III del presente documento, Guha et al (2022, p. 2) sugieren el uso de la metodología IRAC para la elaboración de un benchmark abierto y colaborativo de razonamiento legal, denominado "Legalbench". Este proyecto busca evaluar la corrección del razonamiento legal en LLM, a través de la generación de tareas a resolver por dichos modelos. Se emplea el método IRAC en tanto es un "marco común" para efectuar razonamiento legal, entendido como el proceso de determinar las condiciones legales que provienen de un set de eventos u ocurrencias, con referencia tanto a casos anteriores como a ley codificada (2022, p. 3).

Yu, Quartey y Schilder (2022, pp. 7-8) señalan que emplear *prompts* de razonamiento legal específico –como los basados en el método IRAC y sus variantes– en combinación con *zero-shot prompts* puede mejorar significativamente la correctitud de las respuestas producidas por los LLM que intentan realizar razonamiento legal. En la misma línea, Servantez, Barrow, Hammond et. al. (2024, p. 6) presentan un modelo de *prompt* denominado Chain of Logic que, basado en el método IRAC, se presenta como superior a modelos alternativos para este tipo de razonamiento.

## ii. Los módulos de nuestro sistema

- Módulo *Issue*

El objetivo principal de este módulo es aplicar la LJP en su versión clásica, cuyo planteamiento consiste en determinar si los hechos del caso constituyen una infracción de la ley. En este contexto, la cuestión planteada trata de determinar si los hechos expuestos en el caso satisfacen los requisitos establecidos en la LRCA para sancionar una práctica colusoria horizontal.

El módulo opera simplemente convirtiendo la narración de los hechos en un formato de pregunta, centrándose específicamente en si estos hechos podrían ser considerados anticompetitivos bajo la norma de competencia. No se realiza ningún análisis ni se toma ninguna determinación en esta fase, simplemente estructura la pregunta que guiará el resto del proceso de evaluación jurídica.

- Módulo *Rule*

El módulo se apoya en dos bases de conocimiento externas esenciales para su funcionamiento. La primera es ‘*expert\_notes.xlsx*’, que contiene dos columnas: ‘*goals*’ y ‘*rule*’. La columna ‘*rule*’ contiene artículos relevantes para resolver controversias de Derecho de la Competencia. La columna ‘*goals*’ contiene los objetivos que deben tenerse en cuenta al interpretar estos artículos.

La segunda base de conocimiento comprende las resoluciones emitidas por INDECOPI y OSIPTEL en materia de Derecho de la Competencia. Estas resoluciones han sido descargadas y procesadas por los autores mediante scripts automatizados en Python y librerías de código abierto. La base de datos incluye metadatos de las decisiones, contenido de texto completo y su representación en *embeddings*, proporcionando un repositorio completo de jurisprudencia.

Ambos *datasets* han sido cuidadosamente curados para garantizar su calidad y pertinencia, así como que son de libre acceso, lo que permite al sistema utilizar información actualizada y verificable al analizar conductas contrarias a la competencia.

En cuanto al funcionamiento del módulo, primero realiza una consulta ampliada para mejorar la recuperación de información. Como fue mencionado, la base de conocimientos externa, ‘*expert\_notes.xlsx*’, consta de dos columnas: ‘*goals*’ y ‘*rule*’. La columna ‘*rule*’ contiene los artículos pertinentes, mientras que la columna ‘*goals*’ contiene objetivos que deben tenerse en cuenta al interpretarlos. Esta información se combina posteriormente con los hechos del caso para servir de *input* a un modelo de lenguaje. El LLM genera dos preguntas abstractas relacionadas con la aplicación de cada artículo pertinente a los hechos del caso.

A través de este proceso, se generan preguntas específicas que sirven para dos propósitos. Primero, sirven como *input* para recuperar páginas de decisiones de la base de datos externa que contiene decisiones sobre Derecho de la Competencia. Una vez recuperada esta información, un LLM resume el

contenido extraído. Segundo, sirven como input para que otro LLM explique los criterios de decisiones anteriores de acuerdo con cada pregunta.

Por último, se generan dos archivos: 'metadata\_jurisprudencia.xlsx', que contiene los metadatos de las decisiones utilizadas por nuestro sistema y permite su comparación, y 'jurisprudencia\_relevante.docx', que contiene los resúmenes y las respuestas a las preguntas concretas.

A modo de ejemplo, veamos como el modelo genera los criterios relevantes para interpretar el artículo 4 de la LRCA en el Caso 8:

**Jurisprudencia Relevante**

Las siguientes preguntas: '¿Cuándo una práctica anticompetitiva acordada fuera del territorio nacional puede considerarse que produce efectos dentro del mercado peruano desde una perspectiva geográfica? ¿Bajo qué criterios geográficos se puede determinar que un acuerdo comercial realizado en el extranjero tiene la capacidad de afectar la competencia dentro del territorio peruano?'

sirven para interpretar el siguiente artículo: 'artículo 4.- ámbito de aplicación territorial la presente ley es de aplicación a las conductas que produzcan o puedan producir efectos anticompetitivos en todo o en parte del territorio nacional, aun cuando dicho acto se haya originado en el extranjero.'

El siguiente texto contiene las respuestas a las preguntas: 'Basado en la jurisprudencia proporcionada, puedo responder:

1. Una práctica anticompetitiva acordada fuera del territorio nacional puede considerarse que produce efectos dentro del mercado peruano cuando:

- Afecta a operaciones comerciales que tienen como destino puertos peruanos, aunque las coordinaciones se hayan realizado en diferentes áreas geográficas (p.83-86, Resolución 0171-2020/SDC-INDECOPI).
- La conducta tiene un alcance global que incluye efectos directos en el territorio nacional, independientemente de dónde se originó el acuerdo (p.83-86, Resolución 0171-2020/SDC-INDECOPI).

1) Las dos preguntas abstractas generadas

2) El artículo que debe ser interpretado

3) Los criterios de interpretación basados en decisiones de las agencias de competencia peruanas

- **Módulo Application**

Este módulo tiene como objetivo aplicar los artículos de la LRCA, junto con la interpretación de los mismos, obtenida en el módulo anterior, a los hechos del caso. El primer algoritmo responsable de esta tarea es un LLM con un prompt CoT, contenido en la función 'application()', que aplica la normativa a los hechos del caso, generando el razonamiento y sus conclusiones. Posteriormente, este módulo agrupa las respuestas en dos categorías.

El primer grupo contiene artículos relacionados con el ámbito de aplicación de la LRCA. Todas las respuestas se consolidan en un solo *string* y se aplica de nuevo la función 'application()'. El segundo grupo comprende las respuestas relacionadas con la aplicación de las prácticas colusorias horizontales como infracción. En este caso, se optó por desarrollar otra función, denominada 'practicas\_colusorias()', con un LLM utilizando un prompt CoT específicamente adaptado para desarrollar razonamientos jurídicos basados en el Derecho de la Competencia peruano.

Finalmente, se utiliza otra función llamada 'ira\_resumen()' para resumir las conclusiones en una única salida. El resultado intermedio se almacena en el archivo 'reglas\_aplicacion.docx'.

Al igual que en la sección previa, veamos un *output* del sistema. Nos enfocaremos en el resultado generado por la función 'practicas\_colusorias()' para el Caso 8:

El siguiente texto contiene la síntesis de las respuestas sobre la aplicación de los artículos:  
'Análisis:

Argumentos para prohibición absoluta:

- Existe un acuerdo explícito de fijación de precios entre competidores
- Hay evidencia clara de la conducta (reunión y implementación)
- La fijación de precios está listada en el Art. 11.2.a
- No es complementaria a ningún acuerdo lícito

Argumentos para prohibición relativa:

- Es una práctica INTRAMARCA (entre distribuidores de la misma marca)
- El Art. 11.2 específicamente requiere que sea INTER marca
- No crea barreras a la entrada ni afecta la permanencia de competidores
- Existe competencia significativa de talleres multimarca (60% del mercado)

Respuesta:

La conducta debe evaluarse bajo el régimen de prohibición relativa porque:

1. Aunque involucra fijación de precios, es una práctica INTRAMARCA entre distribuidores de AutoLux, no cumpliendo el requisito fundamental del Art. 11.2 que exige que sea INTER marca para aplicar la prohibición absoluta.
2. No genera efectos exclusorios ni crea barreras a la entrada, existiendo competencia significativa de talleres multimarca.
3. Por tanto, debe analizarse bajo el Art. 11.3 evaluando sus efectos concretos en la competencia y el bienestar del consumidor, no siendo aplicable la prohibición absoluta del Art. 11.2.

La clave está en que el carácter INTRAMARCA de la práctica la excluye automáticamente del ámbito de prohibiciones absolutas, independientemente de que involucre fijación de precios.'

- Módulo *Conclusion*

El módulo de conclusiones tiene como objetivo responder a la pregunta formulada en el módulo 'Issue'. Para ello, en la función denominada 'conclusion()', se instruyó a un LLM con un CoT adaptado para el Derecho de la Competencia. Esta función toma como entrada la pregunta generada en el módulo 'Issue', así como el resumen obtenido en el módulo 'Application', para generar como salida si los hechos del caso constituyen una práctica anticompetitiva.

Por último, se solicitó otro LLM utilizando el esquema JSON para estructurar la información. La información se almacena en los ficheros 'irac.docx' y 'predicción\_ljp.xlsx'.

El *output* final del sistema es el siguiente:

<p>Issue</p>	<p>Los siguientes hechos «En enero de 2019, tres distribuidores autorizados de la marca de automóviles «AutoLux” en Lima Metropolitana - Motorex S.A.C., CarService E.I.R.L. y AutoPremium S.A. - sostuvieron una reunión en un restaurante del distrito de San Isidro. Durante el encuentro, los representantes de las tres empresas discutieron la posibilidad de establecer precios similares para los servicios de mantenimiento preventivo de los vehículos AutoLux en sus respectivos talleres. Los distribuidores acordaron verbalmente fijar un rango de precios para el servicio de mantenimiento de 5,000 kilómetros, estableciendo que este debería oscilar entre S/ 450 y S/ 500. El acuerdo se implementó durante tres meses, entre febrero y abril de 2019. Durante este período, los precios de los servicios de mantenimiento en los tres concesionarios efectivamente se mantuvieron dentro del rango acordado. Sin embargo, durante estos meses, el volumen de servicios de mantenimiento realizados por los tres talleres se mantuvo en niveles similares a los meses previos al acuerdo. Además, los usuarios continuaron teniendo la opción de realizar el mantenimiento de sus vehículos en talleres multimarca, los cuales representaban el 60% de los servicios de mantenimiento para vehículos AutoLux en Lima Metropolitana. Los precios en estos talleres multimarca oscilaban entre S/ 300 y S/ 600 durante el mismo período. A partir de mayo de 2019, cada distribuidor volvió a establecer sus precios de manera independiente, retornando a su política comercial individual’ SI cumple O NO cumple con los requisitos que establece la ley de competencia para sancionar una conducta?</p>
<p>Rule Application Conclusion</p>	<p>‘Análisis:</p> <p>1. Ámbito de aplicación:</p> <p>a) Subjetivo (Art. 2):</p> <ul style="list-style-type: none"> <li>- Los tres distribuidores (Motorex, CarService y AutoPremium) son agentes económicos que realizan actividad económica.</li> <li>- Sus representantes participaron en la conducta.</li> <li>- CUMPLE el ámbito subjetivo.</li> </ul> <p>b) Objetivo (Art. 3):</p> <ul style="list-style-type: none"> <li>- No existe norma legal que autorice la conducta.</li> <li>- No aplica ninguna exclusión.</li> <li>- CUMPLE el ámbito objetivo.</li> </ul> <p>c) Territorial (Art. 4):</p> <ul style="list-style-type: none"> <li>- La conducta se realizó en Lima Metropolitana.</li> <li>- Los efectos se produjeron en el mercado peruano.</li> <li>- CUMPLE el ámbito territorial.</li> </ul>

<p>Rule Application Conclusion</p>	<p>2. Tipo de prohibición:</p> <ul style="list-style-type: none"> <li>- Es una práctica de fijación de precios INTRAMARCA (entre distribuidores de la misma marca AutoLux).</li> <li>- No califica como prohibición absoluta del Art. 11.2 que requiere que sea INTER marca.</li> <li>- Debe evaluarse como prohibición relativa bajo Art. 11.3.</li> </ul> <p>3. Análisis de efectos:</p> <p>Efectos negativos:</p> <ul style="list-style-type: none"> <li>- Eliminación de competencia en precios entre los tres distribuidores autorizados</li> <li>- Restricción artificial del rango de precios (S/450-500)</li> </ul> <p>Efectos limitados por:</p> <ul style="list-style-type: none"> <li>- Duración corta (3 meses)</li> <li>- No hubo cambios en volumen de servicios</li> <li>- Existencia de competencia significativa de talleres multimarca (60% del mercado)</li> <li>- Rango de precios acordado dentro del rango normal del mercado (S/300-600)</li> </ul> <p>Respuesta:</p> <p>La conducta NO cumple con los requisitos para ser sancionada porque:</p> <ol style="list-style-type: none"> <li>1. Si bien cumple con los ámbitos de aplicación de la ley, al ser una práctica INTRAMARCA debe evaluarse como prohibición relativa.</li> <li>2. Bajo el análisis de efectos, aunque existe una restricción a la competencia entre distribuidores autorizados, los efectos negativos no son significativos debido a: <ul style="list-style-type: none"> <li>- La fuerte competencia de talleres multimarca (60% del mercado)</li> <li>- El rango de precios acordado está dentro del rango normal del mercado</li> <li>- No se evidencian efectos en el volumen de servicios</li> <li>- La corta duración de la conducta (3 meses)</li> </ul> </li> </ol> <p>Por tanto, los efectos negativos no son lo suficientemente significativos para justificar una sanción bajo el régimen de prohibición relativa.'</p>
--	---

## i. Resultados

Los resultados de nuestra investigación se dividen en dos aspectos fundamentales: la capacidad explicativa del sistema y su rendimiento predictivo en comparación con modelos de última generación.

### - Explicabilidad del Sistema:

El objetivo principal de nuestro desarrollo fue crear un sistema de IA que pudiera proporcionar explicaciones claras y comprensibles de su proceso de razonamiento en el contexto legal. A diferencia de otros sistemas, nuestro enfoque prioriza la transparencia del proceso de toma de decisiones por encima de la optimización del rendimiento. El sistema genera explicaciones detalladas de su razonamiento mediante la combinación de CoT y RAG, permitiendo seguir paso a paso cómo se procesa la información y se llega a las conclusiones.

Es importante señalar que, en el campo de la explicabilidad de IA para aplicaciones legales, no existe actualmente un estándar establecido o un conjunto de métricas universalmente aceptadas para evaluar la calidad de las explicaciones. Por lo tanto, dejamos a disposición de la comunidad académica y profesional la evaluación cualitativa de nuestras explicaciones, fomentando un diálogo abierto sobre qué constituye una explicación apropiada y útil en este contexto.

### - Rendimiento Predictivo:

Aunque la explicabilidad fue nuestro objetivo primordial, realizamos una evaluación comparativa del rendimiento predictivo de nuestro sistema contra modelos de vanguardia, específicamente los modelos más avanzados de Anthropic y OpenAI. Los resultados fueron notablemente positivos, demostrando que nuestro sistema no solo cumple con el objetivo de ser explicable, sino que también supera en precisión a estos modelos de referencia en la tarea principal de predicción.

Nuestros resultados sugieren que un sistema diseñado con un enfoque en la explicabilidad, cuando se implementa con una arquitectura apropiada (i.e., si esta genera información intermedia que pueda ser entendida por el usuario) y se alimenta con información específica del dominio (como una base de datos de decisiones de una jurisdicción específica), puede superar incluso a modelos más grandes y computacionalmente más poderosos.

claude-3-opus-20240229	o1-preview	ulises
90%	90%	100%

Este resultado puede atribuirse a dos factores principales:

- La integración efectiva de conocimiento específico del dominio legal a través del sistema RAG
- La implementación de *prompts* especializados y optimizados para tareas legales específicas

Estos hallazgos sugieren que la especialización y la explicabilidad pueden ser ventajas competitivas significativas, incluso cuando se compara con modelos de lenguaje más grandes y generales. El programa para hacer esta comparación también se encuentra en el repositorio de GitHub, por lo que fomentamos que se descargue y se reproduzcan los resultados aquí expuestos.

Sobre el particular, los experimentos se realizaron con 11 casos contenidos en el archivo 'dataset\_test.xlsx'. El 28 de diciembre de 2024, se llevó a cabo el experimento utilizando la misma configuración del sistema para cada uno de los casos. Los resultados obtenidos de o1-preview y claude-3-opus-20240229 fueron generados manualmente a través de la consola de la API de OpenAI y Anthropic<sup>16</sup>, respectivamente, y se almacenaron en el documento 'benchmark\_advanced\_models.xlsx'. El detalle del input exacto utilizado se encuentra disponible en el archivo 'input\_for\_advanced\_models.xlsx'.

## V. CONCLUSIONES

Esta investigación ha logrado varios hallazgos significativos en el campo de la XAI aplicada al sector legal. En primer lugar, nuestra revisión exhaustiva de la literatura reveló una predominancia marcada de enfoques cuantitativos en la explicabilidad de modelos de IA, dejando un vacío significativo en áreas donde los métodos numéricos no son predominantes, como es el caso del derecho.

Para abordar esta brecha, hemos implementado y validado métodos alternativos no cuantitativos, específicamente el CoT y el RAG. Estos enfoques han demostrado ser efectivos para hacer transparente el proceso de razonamiento de la IA, permitiendo un seguimiento detallado de cómo el modelo procesa la información y genera sus respuestas. La implementación de resultados intermedios en formatos accesibles a través de software comercial común representa un avance significativo en la practicidad y accesibilidad de estas explicaciones.

Un hallazgo particularmente notable emerge de la comparación de nuestro sistema con el modelo Claude Opus de Anthropic y o1-preview de OpenAI. Utilizando un *benchmark* desarrollado específicamente para esta investigación, nuestro sistema no solo demostró ser más explicable, sino que también superó en precisión a modelos con mayor número de parámetros. Este resultado subraya la importancia de dos factores clave: la alimentación de información externa relevante y el diseño de prompts específicos y bien dirigidos. Estos elementos pueden compensar efectivamente la diferencia en capacidad computacional y tamaño del modelo.

Es muy pronto para decir que llegamos a Ítaca. Sin embargo, confiamos en que, como se sostuvo acertadamente en algún momento<sup>17</sup>, la mejor forma de emprender este viaje es el camino largo, que se nutre de la experiencia. En ese sentido, "Ulises", el sistema resultado de este proyecto es de acceso libre, con la finalidad de permitir su interpelación y este proyecto pueda seguir nutriéndose de la impresión práctica de sus usuarios finales, en este caso, los especialistas en libre competencia.

## VI. FUTURAS DIRECCIONES DE INVESTIGACIÓN

A partir de los hallazgos obtenidos, se identifican múltiples oportunidades para expandir y consolidar la contribución de esta investigación al campo de la XAI en el ámbito legal. En este contexto, se proponen las siguientes líneas de investigación como una hoja de ruta para orientar futuros estudios y avanzar en el desarrollo de soluciones que garanticen un uso responsable de la IA en el sector legal.

- **Exploración de métodos cualitativos en XAI para el ámbito legal:**

---

<sup>16</sup> En ambos casos se utilizó la temperatura 0 como parámetro.

<sup>17</sup> Ver el poema "Ítaca" (1911) del escritor griego Konstantinos Kavafis.

Dado el vacío identificado en el uso de enfoques no cuantitativos en la explicabilidad de la IA, futuras investigaciones podrían centrarse en profundizar y diversificar métodos cualitativos.

- **Optimización de estrategias de alimentación de información externa:**  
Investigar cómo mejorar la selección, integración y actualización de información externa, sea jurisprudencia o artículos científicos, para maximizar la precisión y explicabilidad.
- **Desarrollo de benchmarks especializados en XAI legal:**  
Diseñar benchmarks adicionales para medir la explicabilidad y desempeño de sistemas de IA en tareas legales específicas, fomentando la estandarización y la comparabilidad entre diferentes enfoques y tecnologías.
- **Expansión a otros dominios jurídicos:**  
Extender la investigación a otras áreas del Derecho, como el penal, laboral o administrativo, para evaluar la eficacia y adaptabilidad de los enfoques implementados en contextos legales diversos y con distintas necesidades de explicabilidad.

## REFERENCIAS:

- Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M., Confalonieri, R., Guidotti, R., Del Ser, J., Díaz-Rodríguez, N., & Herrera, F. (2023). Explainable artificial intelligence (XAI): What we know and what is left to attain trustworthy artificial intelligence. *Information Fusion*, 99, 101805. <https://doi.org/10.1016/j.inffus.2023.101805>
- Ash, E., & Chen, D. L. (2018). Case vectors: Spatial representations of the law using document embeddings. In M. Livermore & D. Rockmore (Eds.), *Law as data* (2019(11)). Santa Fe Institute Press. Available at SSRN: <https://ssrn.com/abstract=3204926>
- Ash, E., Jacobs, J., MacLeod, B., Naidu, S., & Stambach, D. (2020). Unsupervised extraction of workplace rights and duties from collective bargaining agreements. In 2020 *International Conference on Data Mining Workshops (ICDMW)* (pp. 766-774). IEEE. <https://doi.org/10.1109/ICDMW51313.2020.00112>
- Baker, S., & Xiang, W. (2023). Explainable AI is responsible AI: How explainability creates trustworthy and socially responsible artificial intelligence. arXiv. <https://arxiv.org/abs/2312.01555>
- Bambroo, P., & Awasthi, A. (2021). LegalDB: Long DistilBERT for legal document classification. In 2021 *International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)* (pp. 1-4). IEEE. <https://doi.org/10.1109/ICAECT49130.2021.9392558>
- Bench-Capon, T. (2020). Explaining legal decisions using IRAC. In Proceedings of CMNA 2020 (pp. 74-83). CEUR Workshop Proceedings, Vol. 2669. Disponible en: <https://ceur-ws.org/Vol-2669/paper10.pdf>
- Bhattacharya, A. (2024). Towards directive explanations: Crafting explainable AI systems for actionable human-AI interactions. *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 1-6. <https://doi.org/10.1145/3613905.3638177>
- Borda, W. (2020). "Alcance e interrelación de las prácticas colusorias horizontales en el marco de la Ley de Represión de Conductas Anticompetitivas". *Themis* (78), pp. 189-206.
- Bullard, A. & Valencia-Dongo, A. (2022). *Competition Law in Peru* [Second Edition – Kindle Version]. Kluwer Law International.
- Chalkidis, I., Androutsopoulos, I., & Aletras, N. (2019). Neural legal judgment prediction in English. In A. Korhonen, D. Traum, & L. Màrquez (Eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 4317-4323). Association for Computational Linguistics. <https://doi.org/10.18653/v1/P19-1424>
- Chalkidis, I., & Kampas, D. (2019). Deep learning in law: Early adaptation and legal word embeddings trained on large corpora. *Artificial Intelligence and Law*, 27(2), 171-198. <https://doi.org/10.1007/s10506-018-9238-9>
- Chalkidis, I., Jana, A., Hartung, D., Bommarito, M., Androutsopoulos, I., Katz, D. M., & Aletras, N. (2021). LexGLUE: A benchmark dataset for legal language understanding in English. *arXiv preprint arXiv:2110.00976*. <https://arxiv.org/abs/2110.00976>
- Columbia Law School I (s/f). Organizing a legal discussion I (IRAC, CRAC, etc.). Columbia Law School Writing Centre. Disponible en: [https://www.law.columbia.edu/sites/default/files/2021-07/organizing\\_a\\_legal\\_discussion.pdf](https://www.law.columbia.edu/sites/default/files/2021-07/organizing_a_legal_discussion.pdf)
- Columbia Law School II (s/f). Organizing a legal discussion II (IRAC, CRAC, etc.). Columbia Law School Writing Centre. Disponible en: <https://www.law.columbia.edu/sites/default/files/2022-06/WC%20Handout%20>

[IRAC%2C%20CRAC%2C%20CREAC.revised%205.22.pdf](#)

Dobson, N. (2020). 2-page Guide Series: Skill in the spotlight: IRAC method for answering questions. Utrecht University School of Law. Disponible en <http://www.uu.nl/sites/default/files/rebo-LSA-IRAC%20method%20final.pdf>

Falla, A. (2019). *Política de Competencia en el Perú: Hay mucho pan por rebanar* [Publicado el 10 de diciembre de 2019]. Centro de Competencia de la Universidad Adolfo Ibañez. Disponible en [www.centrocompetencia.com/politica-de-competencia-en-el-peru-hay-mucho-pan-por-rebanar](http://www.centrocompetencia.com/politica-de-competencia-en-el-peru-hay-mucho-pan-por-rebanar)

Feng, Y., Li, C., & Ng, V. (2022). Legal judgment prediction: A survey of the state of the art. In L. De Raedt (Ed.), *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22* (pp. 5461–5469). International Joint Conferences on Artificial Intelligence Organization. <https://doi.org/10.24963/ijcai.2022/765>

Guha, N., et al. (2023). LegalBench: A collaboratively built benchmark for measuring legal reasoning in large language models. *arXiv preprint arXiv:2308.11462*. <https://arxiv.org/abs/2308.11462>

Guo, X., & Chen, Y. (2024). Generative AI for synthetic data generation: Methods, challenges and the future. arXiv. <https://arxiv.org/abs/2403.04190>

Han, Z., & Dou, Z. (2023). Case retrieval for legal judgment prediction in legal artificial intelligence. In M. Sun, B. Qin, X. Qiu, J. Jiang, & X. Han (Eds.), *Proceedings of the 22nd Chinese National Conference on Computational Linguistics* (pp. 801–812). Chinese Information Processing Society of China. <https://aclanthology.org/2023.ccl-1.68>

Kang, X., Qu, L., Soon, L. K., Li, Z., & Trakic, A. (2024). Bridging Law and Data: Augmenting Reasoning via a Semi-Structured Dataset with IRAC methodology. Disponible en: <https://arxiv.org/pdf/2406.13217>

Kang, X., Qu, L., Soon, L. K., Trakic, A., Zhuo, T. Y., Emerton, P. C., & Grant, G. (2023). Can ChatGPT Perform Reasoning Using the IRAC Method in Analyzing Legal Scenarios Like a Lawyer? Disponible en: <https://arxiv.org/abs/2310.14880>

Katz, D. M., Hartung, D., Gerlach, L., Jana, A., & Bommarito, M. J., II. (2023). *Natural language processing in the legal domain*. arXiv. <https://arxiv.org/abs/2302.12039>

Luz de Araujo, P. H., de Campos, T. E., Ataide Braz, F., & Correia da Silva, N. (2020). VICTOR: A dataset for Brazilian legal documents classification. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 1449–1458). European Language Resources Association. <https://aclanthology.org/2020.lrec-1.181>

Maruthi, S., Babu Dodda, S., Reddy Yellu, R., Thuniki, P., & Reddy Byrapu Reddy, S. (2022). Language Model Interpretability - Explainable AI Methods: Exploring explainable AI methods for interpreting and explaining the decisions made by language models to enhance transparency and trustworthiness. *Australian Journal of Machine Learning Research & Applications*, 2(2), 1-9. <https://sydneyacademics.com/index.php/ajmlra/article/view/19>

Medvedeva, M., & McBride, P. (2023). Legal judgment prediction: If you are going to do it, do it right. In D. Preotescu-Pietro, C. Goanta, I. Chalkidis, L. Barrett, G. Spanakis, & N. Aletras (Eds.), *Proceedings of the Natural Legal Language Processing Workshop 2023* (pp. 73–84). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.nllp-1.9>

Meller, G. (2024). *Explainable artificial intelligence: A study of methods, applications, and future directions* [Master's

thesis, Alma Mater Studiorum Università di Bologna, Department of Computer Science and Engineering]. [https://amslaurea.unibo.it/id/eprint/32790/1/XAI\\_Master\\_Thesis.pdf](https://amslaurea.unibo.it/id/eprint/32790/1/XAI_Master_Thesis.pdf)

Munn, M., & Pitman, D. (2023). *Explainable AI for practitioners: Designing and implementing explainable ML solutions*. O'Reilly Media.

Patrón, C. (2008). "Aciertos, divergencias y desatinos de la nueva Ley de Represión de Conductas Anticompetitivas". *Ius Et Veritas* 18(36), pp. 122-144.

Srihari, S. N. (2024). *Overview of Explainable Artificial Intelligence (XAI)*. Department of Computer Science and Engineering, University at Buffalo, The State University of New York. Retrieved from <https://cedar.buffalo.edu/~srihari/papers/XAI-Overview.pdf>

OECD (2018). *Exámenes Inter-Pares de la OCDE y el BID sobre el Derecho y Política de Competencia: Perú*. Paris: OECD Publishing. Disponible en: <https://doi.org/10.1787/af0ef152-es>

OECD. (2023). *Advancing accountability in AI: Governing and managing risks throughout the lifecycle for trustworthy AI* (OECD Digital Economy Papers, No. 349). OECD Publishing

Östling, A., Sargeant, H., Xie, H., Bull, L., Terenin, A., Jonsson, L., Magnusson, M., & Steffek, F. (2023). The Cambridge Law Corpus: A dataset for legal AI research. *arXiv preprint arXiv:2309.12269*. <https://arxiv.org/abs/2309.12269>

Peng, X., & Chen, L. (2024). Athena: Retrieval-augmented legal judgment prediction with large language models. *arXiv*. <https://arxiv.org/abs/2410.11195>

Quintana, E. (2013). *Libre Competencia*. Lima: Instituto Nacional de Defensa de la Competencia y de la Protección de la Propiedad Intelectual.

Raji, I. D., Bender, E. M., Paullada, A., Denton, E., & Hanna, A. (2021). AI and the everything in the whole wide world benchmark. *arXiv preprint arXiv:2111.15366*. <https://arxiv.org/abs/2111.15366>

Servantez, S., Barrow, J., Hammond, K., & Jain, R. (2024). Chain of Logic: Rule-Based Reasoning with Large Language Models. Disponible en: <https://arxiv.org/pdf/2402.10400>

Sherkow, J. S. (2024). Putting the C Back in IRAC. *Green Bag.2d*, 27, 39-49. University of Illinois College of Law Legal Studies Research Paper No. 24-12. <https://ssrn.com/abstract=4734676>

Sulea, O.-M., Zampieri, M., Malmasi, S., Vela, M., Dinu, L. P., & van Genabith, J. (2017). Exploring the use of text classification in the legal domain. *arXiv preprint arXiv:1710.09306*. <https://arxiv.org/abs/1710.09306>

Tuggener, D., von Däniken, P., Peetz, T., & Cieliebak, M. (2020). LEDGAR: A large-scale multi-label corpus for text classification of legal provisions in contracts. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 1235-1241). European Language Resources Association. <https://aclanthology.org/2020.lrec-1.155>

Vainio-Pekka, H., Agbese, M. O., Jantunen, M., Vakkuri, V., Mikkonen, T., Rousi, R., & Abrahamsson, P. (2023). The role of explainable AI in the research field of AI ethics. *ACM Transactions on Interactive Intelligent Systems*, 13(4), Article 26, 39 pages. <https://doi.org/10.1145/3599974>

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2023). Chain-of-thought prompting elicits reasoning in large language models. *arXiv*. <https://arxiv.org/abs/2201.11903>





Este documento se encuentra sujeto a los términos y condiciones de uso disponibles en nuestro sitio web:  
<http://www.centrocompetencia.com/terminos-y-condiciones/>

**Cómo citar este artículo:**

Piero Alexis Malca Vilchez, Enzo Rodrigo Gómez Rojas y César Humberto Quiñones Costa, "El viaje de Ulises: Hacia un modelo de inteligencia artificial explicable para resolver problemas de libre competencia",  
*Investigaciones CeCo* (marzo, 2025),

<http://www.centrocompetencia.com/category/investigaciones>

Envíanos tus comentarios y sugerencias a [info@centrocompetencia.com](mailto:info@centrocompetencia.com)  
CentroCompetencia UAI – Av. Presidente Errázuriz 3485, Las Condes, Santiago de Chile