



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERIA INDUSTRIAL

DESARROLLO DE UN MODELO DE PROCESAMIENTO DE LENGUAJE NATURAL PARA LA INTERPRETACIÓN JUDICIAL DE CONTRATOS

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

ISMAEL ANDRÉS ORREGO VENEGAS

PROFESOR GUÍA:
JUAN VELASQUEZ SILVA

MIEMBROS DE LA COMISIÓN:

JUAN PABLO ROMERO GODOY
RODRIGO COLOMA CORREA

Este trabajo ha sido parcialmente financiado por Fondecyt Regular N°1210127

SANTIAGO DE CHILE

2023

DESARROLLO DE UN MODELO DE PROCESAMIENTO DE LENGUAJE NATURAL PARA LA INTERPRETACIÓN JUDICIAL DE CONTRATOS

El objetivo general de la memoria es el diseño y construcción de un modelo basado en Procesamiento de Lenguaje Natural (NLP) que permita analizar los textos de demandas y sentencias, para la extracción de patrones, que a partir de lo estipulado en los artículos 1560 a 1566 del Código Civil, contribuya a la producción de soluciones estandarizadas para profesionales del derecho en Chile.

Dada la ambigüedad presente en el Código Civil chileno para la interpretación de contratos, se evidencia una multiplicidad de perspectivas que derivan en desacuerdos significativos entre tribunales. Esta realidad, no solo desgasta la confianza en el sistema judicial, sino que también agota recursos valiosos. Esta problemática subraya la imperante necesidad de herramientas precisas y estandarizadas para enfrentar estos desafíos interpretativos.

Ante el panorama descrito, emerge la hipótesis de que es posible utilizar la tecnología actual para entrenar un modelo que genere respuestas coherentes y alineadas con las decisiones de la Corte Suprema respecto a la interpretación judicial de contratos.

Mientras que el escenario tecnológico actual muestra avances significativos, con empresas privadas lanzando asistentes judiciales digitales como "Sofía de Tirant", la iniciativa pública en este ámbito parece estancada. Estos desarrollos privados son testigos de la capacidad de los algoritmos para desentrañar el lenguaje jurídico, pero la falta de alternativas o desarrollo de carácter público amplifica la relevancia de este estudio.

Siguiendo la metodología CRISP-DM, se determinó que los Modelos de Lenguaje a Gran Escala (LLM) son idóneos para este propósito. El reciente modelo gpt-3.5-turbo de OpenAI, liberado para el ajuste fino, destacó en la investigación. Evaluando e iterando sobre el nivel de respuesta de las herramientas disponibles, se creó una muestra de preguntas y respuestas para realizar el ajuste fino, a pesar de entrenarse con una muestra significativamente menor a la recomendada, demostró producir resultados sorprendentemente cercanos a las decisiones de la Corte Suprema.

Concluyendo, obtener semejantes resultados con un presupuesto y una muestra particularmente acotada, indican un horizonte prometedor para futuras investigaciones. Los algoritmos en desarrollo poseen el potencial para transformar radicalmente la interpretación de contratos en el sistema judicial chileno, aspirando a respuestas más pulidas, estandarizadas y, en última instancia, a un sistema más equitativo.

Tabla de Contenido

1	Introducción	1
1.1	Descripción del Problema	2
1.2	Descripción del Proyecto	2
1.3	Hipótesis	3
1.4	Objetivos	3
1.4.1	Objetivo General.....	3
1.4.2	Objetivos Específicos	3
1.5	Metodología.....	4
1.6	Contribuciones de la Memoria	5
1.7	Estructura de la Memoria	5
2	Marco Conceptual	6
2.1	Código Civil	7
2.1.1	Libro IV: De las obligaciones en general y de los contratos	7
2.2	Código Orgánico De Tribunales.....	8
2.2.1	Título IX De los Jueces Árbitros.....	9
2.3	Natural Language Processing (NLP)	9
2.3.1	Conceptos introductorios	10
2.3.2	Estado del Arte	14
2.3.3	Legal Tech	20
2.3.4	Herramientas para el proyecto	21
3	Levantamiento de Requerimientos	24
4	Diseño y Construcción de la Solución	28
4.1	Comprensión de los datos	28
4.2	Preparación de los datos	39
4.3	Modelado	40
4.4	Evaluación	40
5	Análisis de Resultados.....	45
5.1	Alternativas desechadas	45
5.2	Implicancias de los resultados	46
5.3	Recomendaciones para el futuro	47
6	Conclusiones.....	48
7	Bibliografía	50

Capítulo 1

Introducción

En los últimos años, el procesamiento de lenguaje natural (PLN o NLP por sus siglas en inglés) ha experimentado notables avances, lo que ha impulsado la exploración de sus aplicaciones en diversas áreas, desde la minería de datos hasta el análisis de sentimientos, y más recientemente, en el mundo jurídico, con la interpretación de textos legales como políticas, leyes y contratos. Sin embargo, en Chile, no se ha realizado ningún avance significativo ni aplicación de estas tecnologías emergentes en el ámbito legal. Por ello, surge la idea de crear una primera aplicación real de estas tecnologías para este ámbito, iniciando así el desarrollo e investigación de estas herramientas en el país. El enfoque de esta memoria se centrará en la aplicación de estas tecnologías para la interpretación judicial de contratos, facilitando las tareas de los actores involucrados en el Poder Judicial, con relación a la resolución de juicios asociados a contratos.

El Poder Judicial en Chile constituye uno de los tres pilares que sostienen el estado democrático de derecho, junto al Poder Ejecutivo y Legislativo. Su misión es otorgar a las personas una justicia oportuna y de calidad. La Corte Suprema está a la cabeza de este poder, y bajo ella se sitúan 17 Cortes de Apelaciones distribuidas en las distintas regiones del país. Dependientes de estas últimas, se encuentran 465 tribunales de primera instancia distribuidos en todo el territorio nacional. Los tribunales de primera instancia son el punto de partida en los asuntos jurídicos y contemplan variados juzgados: Familia, Cobranza Laboral y Previsional, Letras del Trabajo, Civiles, entre otros. [1]

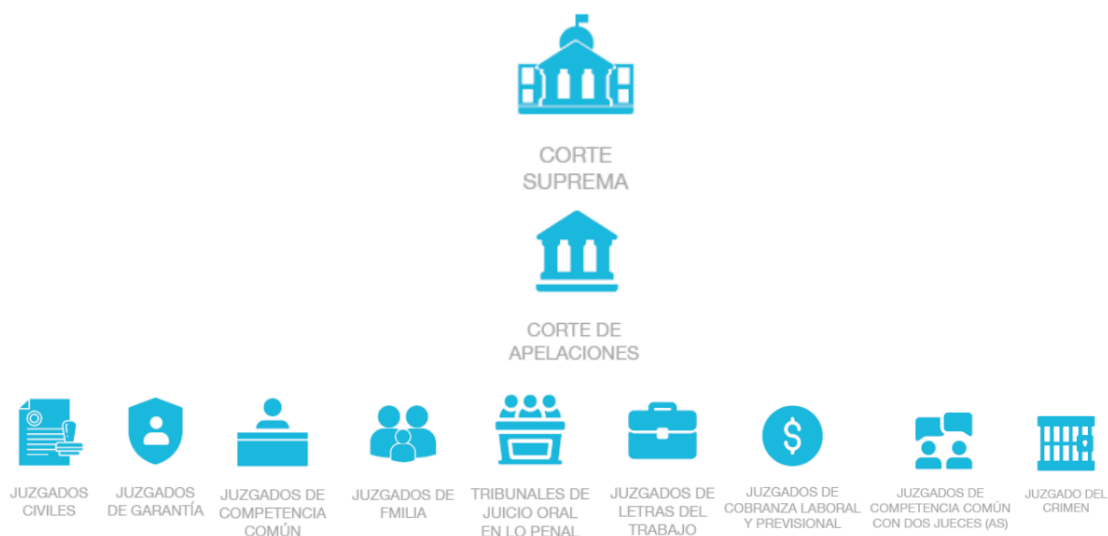


Ilustración 1: Estructura del Poder Judicial

1.1 Descripción del Problema

Los contratos regulan los derechos y obligaciones que benefician o recaen entre dos o más partes que participan en un negocio o intercambio de bienes. En principio, existe la expectativa que lo que en el contrato se diga sea suficientemente claro como para resolver cualquier desacuerdo futuro entre las partes durante su ejecución. A pesar de que en la mayoría de las ocasiones no se producen tales desacuerdos, hay otras en que así ocurre y aquello puede ser planteado ante los tribunales de justicia para que lo decidan.

El Código Civil tiene varias disposiciones que tratan de la interpretación de contratos. Estas disposiciones debieran guiar el proceso de toma de decisión de los tribunales de justicia y, de esta manera, debieran resultar medianamente previsibles y, a la vez, concordantes, cuando se trata de casos parecidos. El conjunto de disposiciones no es completo. Las aludidas disposiciones sobre interpretación de contratos del Código Civil admiten múltiples significados y formas de organización del material a interpretar. Es preciso, entonces, elegir entre dichas posibilidades.

A la fecha, hay un nivel importante de desacuerdos entre los tribunales de justicia respecto a la manera en que deben resolverse los asuntos que conocen sobre interpretación de contratos. No es infrecuente que lo que un tribunal inferior dijo, sea modificado por el superior e, incluso, que entre los distintos integrantes de estos últimos haya desacuerdos. Los contratos son muy diversos, por una parte, un mismo propósito se puede plantear de maneras muy diferentes; y por la otra, dos textos parecidos podrían servir para comunicar asuntos bastante distintos.

Es por esto que es difícil que la situación se regularice por sí sola, puesto que actualmente las disposiciones brindan el espacio para la libre interpretación, dejando sin un marco claro con el que regirse, y permitiendo que cada quien defienda la interpretación que más le convenga.

1.2 Descripción del Proyecto

En base a lo mencionado anteriormente, el abogado y profesor titular de derecho, Rodrigo Coloma Correa, quien ya había escrito un artículo de doctrina sobre la falta de operatividad de estas disposiciones [2], postula junto a su equipo, a un proyecto FONDECYT (Fondo Nacional de Desarrollo Científico y Tecnológico), consiguiendo la aprobación para llevar a cabo una investigación acerca del tema.

Dentro de este contexto es que se hizo el espacio dentro de la misma investigación, para investigar sobre la posibilidad de encontrar algoritmos de inteligencia artificial, que permitan contribuir a la producción de soluciones, que permitan guiar a los abogados en torno a la redacción de cláusulas contractuales, o bien a los jueces para resolver como dirimir conflictos sobre interpretación de contratos en concreto.

Por consiguiente, que se vuelve un acercamiento de interés, el explorar si es posible la aplicación de la ciencia computacional, para contribuir a la producción de soluciones en base a los patrones que se puedan encontrar. La memoria entonces toma lugar dentro del contexto de un Proyecto FONDECYT, de investigación científica aplicada, en torno a la potencial utilización de algoritmos de Machine Learning para la interpretación judicial de contratos.

En principio se vislumbraban dos posibles alternativas de solución:

Una que en base a las cláusulas más controversiales encontradas o reconocidas en las diversas sentencias de los contratos, formule un tipo de punteo o de guía para los abogados, en torno a la redacción de estas cláusulas contractuales, de modo de aclarar o de completar, las partes que se encuentran abiertas a interpretación dentro del contrato, lo que potencialmente evitaría o disminuiría de entrada, la llegada de juicios asociados a interpretación de estos.

Otra solución, es una que, a partir del resumen del conflicto de una situación en concreto, se formule una respuesta o guía de cómo dirimirse en base a la situación planteada, sirviendo como ayuda para estandarizar soluciones y disminuir los desacuerdos entre los pares de los tribunales, liberando así una cantidad significativa de horas y recursos de los agentes involucrados, y promoviendo la generación de juicios más justos.

El trabajo de esta memoria trata de realizar la segunda alternativa de solución, siendo limitado principalmente por la cantidad, diversidad y tipo de situaciones facilitadas.

1.3 Hipótesis

El trabajo desarrollado en este documento está motivado por las siguientes hipótesis de investigación:

- Es factible con la tecnología actual, rescatar el sentido semántico de oraciones judiciales chilenas de gran complejidad.
- Es factible con la tecnología actual, identificar estructuras tipo de demandas o sentencias chilenas, con herramientas de Machine Learning.
- Es factible con la tecnología actual, entrenar un modelo que genere respuestas similares a la Corte Suprema, a conflictos de interpretación judicial de contratos.

1.4 Objetivos

Para direccionar entonces tanto el desarrollo como el alcance del proyecto, se plantean los siguientes objetivos:

1.4.1 Objetivo General

Diseñar y construir un modelo de interpretación, basado en Natural Language Processing, que analice los contenidos textuales presentes en demandas y sentencias, para la extracción de patrones, que a partir de lo que estipulan los artículos 1560 a 1566 del Código Civil, y de las distintas sentencias, contribuya a la producción de soluciones estandarizadas para abogados y jueces.

1.4.2 Objetivos Específicos

Para lograr el objetivo general, se proponen los siguientes objetivos específicos:

- Desarrollar un estado del arte de algoritmos y técnicas NLP para el análisis de contratos
- Definir y dimensionar los documentos, formatos y magnitudes pertinentes para el modelo.

- Evaluar el nivel de resultados de modelos actuales e identificar posibles mejoras.
- Construir un modelo con la documentación recopilada y evaluar sus resultados.
- Documentar el proceso y dar recomendaciones para potenciales mejoras del modelo.

1.5 Metodología

La metodología escogida para este proyecto fue la de CRISP-DM (Cross-Industry Standard Process for Data Mining), ya que destaca por su flexibilidad, al presentar fases que pueden superponerse o repetirse según sea necesario, estas fases son:

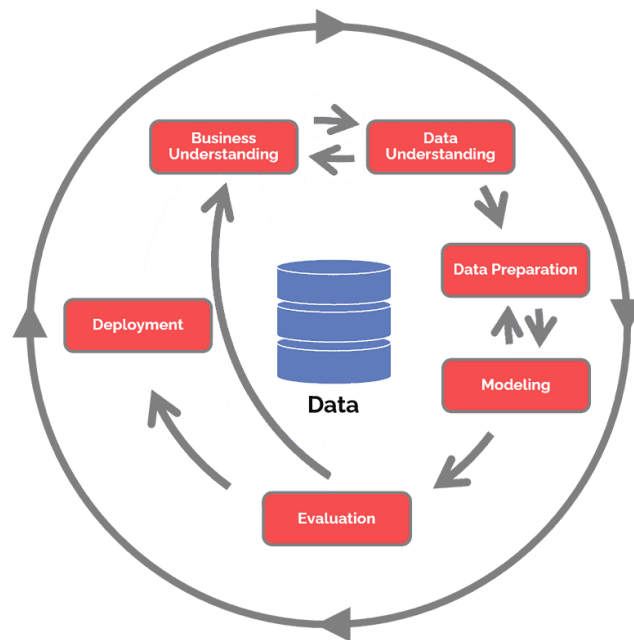


Ilustración 2: Diagrama extraído de "What is CRISP DM?" [3]

1. Comprensión del negocio: Se comprende y se define lo que necesita o desea lograr el cliente.
2. Comprensión de los datos: Se recopilan y se exploran los datos con los que se disponen.
3. Preparación de los datos: Se seleccionan los datos más apropiados, se limpian y se transforman en los formatos adecuados para los modelos.
4. Modelado: Se seleccionan los modelos o algoritmos a utilizar y se calibran sus parámetros.
5. Evaluación: Se evalúa si los resultados cumplen los objetivos definidos.
6. Despliegue: Se entregan los resultados o el producto al cliente.

1.6 Contribuciones de la Memoria

Las contribuciones centrales de este trabajo de memoria se pueden condensar en los siguientes puntos:

- **Innovación Documental:** Este trabajo presenta el primer documento que aborda de manera exhaustiva el alcance, tanto teórico como práctico, de la tecnología vigente. En particular, se enfoca en la aplicación de herramientas de Natural Language Processing (Procesamiento de Lenguaje Natural) en la interpretación judicial chilena. Esto sienta un precedente y proporciona una base avanzada para investigaciones subsecuentes.
- **Metodología Evaluativa:** Se introduce una metodología diseñada para evaluar la calidad de las respuestas generadas por diferentes modelos de NLP. Esta evaluación se realiza a través de consultas concisas y meticulosamente formuladas, lo que puede ser un recurso valioso para futuros trabajos en esta área.
- **Exploración Empírica:** A través de un análisis empírico, se destaca la brecha existente entre la tecnología actual y el sistema de justicia chileno. Este hallazgo subraya el inmenso potencial y la necesidad de continuar investigando la sinergia entre estos dos campos.

1.7 Estructura de la Memoria

El resto de este documento se estructura de la siguiente manera:

En el Capítulo 2 se detalla el marco conceptual en el que se desarrolla el proyecto, detallando lo que es un contrato, cómo funciona nuestro sistema legal, el arbitraje, qué es el Procesamiento de Lenguaje Natural, su estado del arte y que herramientas han de ser consideradas para el proyecto.

El Capítulo 3 se describe el trabajo previo realizado y las iteraciones que fueron necesarias para decantarse por el desarrollo del modelo final.

En el Capítulo 4 se presenta el proceso de diseño y construcción de la solución, es decir, el cómo se hizo, y qué resultados dieron.

En el Capítulo 5 se presentan los análisis finales rescatados a partir de los resultados vistos previamente.

Finalmente, en el Capítulo 6 se presentan las conclusiones derivadas del trabajo desarrollado, y se describen posibles direcciones de trabajo futuro.

Capítulo 2

Marco Conceptual

En Chile, el proceso judicial se rige por el Código de Procedimiento Civil y por otras leyes complementarias. El orden de un juicio en Chile es el siguiente:

Demanda: El proceso comienza cuando una de las partes presenta una demanda ante el tribunal competente, que describe el motivo de la disputa y lo que busca obtener de la otra parte, ésta debe ser patrocinada por un abogado y éste debe contar con un poder que lo faculte para actuar en representación del demandante. La demanda es un documento en el que se establecen los hechos, las pruebas y los fundamentos jurídicos en los que se basa la pretensión del demandante.

Notificación: El tribunal, una vez recibida la demanda, la analiza y, si se cumplen los requisitos legales, ordena la notificación del demandado. La notificación es el acto por el cual se le informa al demandado de la existencia de la demanda y se le da la oportunidad de defenderse.

Contestación: El demandado tiene un plazo para contestar la demanda. La contestación es un documento en el que el demandado responde a los hechos y a las pretensiones del demandante y presenta sus propios argumentos y pruebas. En el proceso ordinario también hay espacio para una réplica y una dúplica.

Periodo de prueba: Una vez presentada la contestación, se convoca a una audiencia preliminar en la que las partes exponen sus argumentos y pruebas y se discuten los puntos en litigio. El objetivo de la audiencia preliminar es simplificar el proceso y definir las pruebas que se admitirán en el juicio.

Sentencia: Una vez que se han presentado todas las pruebas y argumentos, el tribunal dicta una sentencia en la que se resuelve la disputa.

Apelación: Si una parte no está satisfecha con la sentencia dictada por el tribunal, puede presentar una apelación ante la Corte de Apelaciones de la Región. Es importante tener en cuenta que la apelación no suspende la ejecución de la sentencia dictada por el tribunal, a menos que se otorgue una medida cautelar especial. Además, el recurso de apelación se limita a la revisión de la sentencia dictada por el tribunal, por lo que no es posible presentar nuevas pruebas o argumentos.

Corte Suprema: Si una parte no está satisfecha con la sentencia dictada por la Corte de Apelaciones, tiene la posibilidad de presentar un recurso de casación en el fondo o en la forma ante la Corte Suprema. Es importante destacar que el recurso de casación en el fondo o en la forma solo puede ser presentado por cuestiones de derecho y no por cuestiones de hecho. Es decir, no se pueden presentar nuevas pruebas ni argumentos en este recurso, sino que se deben limitar a alegar errores en la aplicación del derecho.

En Chile, existen numerosos códigos legales que cubren diferentes áreas del derecho, a la fecha existen los siguientes 15: Código Aeronáutico, Código Civil, Código Orgánico De Tribunales, Código Penal, Código Procesal Penal, Código Sanitario, Código Tributario, Código De Aguas, Código de Comercio, Código De Derecho Internacional Privado, Código De Justicia Militar, Código De Minería, Código De Procedimiento Civil, Código De Procedimiento Penal, y Código Del Trabajo.

Hay muchas otras leyes y regulaciones en Chile además de estos códigos, y la cantidad exacta de códigos legales puede variar según la interpretación de lo que se considera un "código". La razón por la cual Chile tiene varios códigos es porque el país se rige por el sistema de derecho romano-germánico, que se caracteriza por la existencia de códigos que regulan diferentes áreas del derecho. Esto es diferente al sistema de derecho común, que se basa en la jurisprudencia y la interpretación de las leyes por parte de los tribunales.

En esta memoria serán de interés el Código Civil y el Código Orgánico de Tribunales.

2.1 Código Civil

El Código Civil de la República de Chile es el cuerpo legal que regula sustancialmente las materias jurídicas civiles del país. Fue redactado por Andrés Bello en 1855 y es considerado como uno de los principales códigos civiles de América Latina. En él, se establece las normas en materia de derechos y obligaciones de las personas, el régimen de bienes, las sucesiones, las obligaciones, los contratos, entre otros temas importantes del derecho civil. Además, ha sido modificado en varias ocasiones para adaptarse a los cambios sociales y legales del país.

El Código Civil de la República de Chile se divide en cuatro libros: Libro I: De las personas, Libro II: De los bienes y de su dominio, posesión, uso y goce, Libro III: De la sucesión por causa de muerte y de las donaciones entre vivos, Libro IV: De las obligaciones en general y de los contratos.

2.1.1 Libro IV: De las obligaciones en general y de los contratos

Esta sección es la más extensa del Código Civil, contempla desde el artículo 1437 al 2524, y además es el de particular interés para el tema de esta memoria.

La definición de contrato se encuentra en el artículo 1438: *“Contrato o convención es un acto por el cual una parte se obliga para con otra a dar, hacer o no hacer alguna cosa. Cada parte puede ser una o muchas personas.”* [4]

Más adelante se dan las regulaciones principales para los tipos de contratos más comunes (arrendamiento, compraventa, permuta, etc.). El conflicto ocurre mayormente con los artículos del Título XIII: De la interpretación de los contratos, que contempla desde el artículo 1560 al 1566.

Art. 1560: *“Conocida claramente la intención de los contratantes, debe estarse a ella más que a lo literal de las palabras.”*

Art. 1561: *“Por generales que sean los términos de un contrato, sólo se aplicarán a la materia sobre que se ha contratado.”*

Art. 1562: *“El sentido en que una cláusula puede producir algún efecto, deberá preferirse a aquel en que no sea capaz de producir efecto alguno.”*

Art. 1563: *“En aquellos casos en que no apareciere voluntad contraria deberá estarse a la interpretación que mejor cuadre con la naturaleza del contrato. Las cláusulas de uso común se presumen aunque no se expresen.”*

Art. 1564: *“Las cláusulas de un contrato se interpretarán unas por otras, dándose a cada una el sentido que mejor convenga al contrato en su totalidad. Podrán también interpretarse por las de otro contrato entre las mismas partes y sobre la misma materia. O por la aplicación práctica que hayan hecho de ellas ambas partes, o una de las partes con aprobación de la otra.”*

Art. 1565: *“Cuando en un contrato se ha expresado un caso para explicar la obligación, no se entenderá por sólo eso haberse querido restringir la convención a ese caso, excluyendo los otros a que naturalmente se extienda.”*

Art. 1566: *“No pudiendo aplicarse ninguna de las reglas precedentes de interpretación, se interpretarán las cláusulas ambiguas a favor del deudor. Pero las cláusulas ambiguas que hayan sido extendidas o dictadas por una de las partes, sea acreedora o deudora, se interpretarán contra ella, siempre que la ambigüedad provenga de la falta de una explicación que haya debido darse por ella.”*

Debido a lo difuso que pueden llegar a ser estos artículos, es que se da el espacio para que se le asocien múltiples usos, fácilmente moldeables a las ocasionales preferencias de cada intérprete, lo que da a lugar a que no exista una clara consistencia en las decisiones judiciales asociadas.

2.2 Código Orgánico De Tribunales

El Código Orgánico de Tribunales (COT) es el compendio legal de Chile que regula la estructura y funciones de los tribunales de justicia que forman parte del Poder Judicial, así como el estatuto y funciones de los auxiliares de la administración de justicia. Promulgado en 1943, el COT engloba la mayoría de las disposiciones sobre derecho procesal orgánico en Chile.

El Poder Judicial chileno es uno de los tres pilares fundamentales que sostienen el estado democrático de derecho, junto al Poder Ejecutivo y Legislativo. Su misión es proporcionar a las personas una justicia oportuna y de calidad. Presidida por la Corte Suprema, la estructura judicial incluye 17 Cortes de Apelaciones distribuidas en diferentes regiones del país, y bajo ellas, 465 tribunales de primera instancia repartidos en todo el territorio nacional.

Estos tribunales de primera instancia son el inicio de los procesos jurídicos y abarcan diversos juzgados: de Familia, Cobranza Laboral y Previsional, Letras del Trabajo, Civiles y Letras de Competencia Común. También incluyen tribunales especiales como los tributarios aduaneros, de contratación pública, de propiedad industrial, de libre competencia y ambientales.

2.2.1 Título IX De los Jueces Árbitros

El arbitraje se establece como un método alternativo para la resolución de conflictos, donde las partes involucradas en un contrato deciden someter sus disputas a un tercero imparcial, el árbitro, en vez de acudir a los tribunales ordinarios.

Art. 222: “Se llaman árbitros, los jueces nombrados por las partes, o por la autoridad judicial en subsidio, para la resolución de un asunto litigioso”

Las partes en un contrato pueden tener varias razones para preferir el arbitraje sobre los tribunales ordinarios. Una de las razones principales es la confidencialidad. El arbitraje proporciona un nivel de privacidad y confidencialidad superior al de los tribunales ordinarios, aspecto crucial en casos que involucren información comercial o empresarial sensible.

Otra ventaja importante del arbitraje es la rapidez. Generalmente, el proceso de arbitraje es más ágil que el judicial, ya que se adapta a las necesidades de las partes y puede realizarse sin tener que cumplir con ciertos requisitos procesales formales. Además, la especialización también es un factor clave. Las partes pueden seleccionar un árbitro con experiencia o conocimientos específicos en el área en disputa, lo que puede potenciar la calidad y eficiencia del proceso de resolución de conflictos.

La flexibilidad es otra ventaja significativa del arbitraje. Permite a las partes establecer las reglas y procedimientos que regirán el proceso, lo que puede ser especialmente beneficioso en contratos complejos o internacionales, donde las partes pueden tener diferentes necesidades y preferencias.

En conclusión, el arbitraje se presenta como una opción atractiva para resolver disputas contractuales en Chile debido a su rapidez, confidencialidad, especialización y flexibilidad en comparación con los tribunales ordinarios.

2.3 Natural Language Processing (NLP)

El Procesamiento del Lenguaje Natural (NLP, por sus siglas en inglés) es una rama de la inteligencia artificial que se enfoca en la interacción entre las computadoras y los humanos utilizando lenguaje natural. El objetivo del NLP es hacer que las computadoras sean capaces de comprender, interpretar y generar lenguaje humano de manera similar a como lo hacen los seres humanos.

Este tiene aplicaciones en una amplia variedad de campos, incluyendo la traducción automática, el reconocimiento de voz, la generación de texto, la búsqueda de información, el análisis de sentimientos, entre otros. Por ejemplo, los motores de búsqueda utilizan NLP para identificar y clasificar el contenido relevante en la web, mientras que las redes sociales lo utilizan para analizar el sentimiento de los usuarios hacia una marca o un producto.

El NLP se basa en una serie de técnicas y herramientas que permiten a las computadoras comprender el lenguaje natural. Algunas de las técnicas comunes utilizadas en el NLP incluyen la tokenización, que consiste en dividir un texto en palabras o frases más pequeñas, y la eliminación de stop words, que son palabras comunes como "el" y "de" que normalmente no aportan mucho significado al texto.

Otras técnicas utilizadas en el NLP incluyen el análisis sintáctico, que se utiliza para analizar la estructura gramatical de una oración, y el análisis semántico, que se utiliza para comprender el significado de las palabras y frases en un texto. También se utilizan técnicas de aprendizaje automático, que permiten a las computadoras aprender de los datos para mejorar la precisión y eficacia del análisis.

El Procesamiento del Lenguaje Natural es una disciplina que tiene un gran potencial para cambiar la forma en que las personas interactúan con las computadoras y la tecnología en general. A medida que la tecnología continúa avanzando, es probable que desempeñe un papel cada vez más importante en la forma en que las personas interactúan con el mundo digital que les rodea.

2.3.1 Conceptos introductorios

Sentence Segmentation: La segmentación de oraciones consiste en dividir un texto en oraciones individuales. La segmentación de oraciones es importante porque muchas aplicaciones, como la traducción automática y la generación de resúmenes, requieren una comprensión precisa de las oraciones del texto.

Existen varios enfoques para la segmentación de oraciones, desde la regla simple de dividir el texto en función de los signos de puntuación, hasta el uso de modelos de aprendizaje automático avanzados que reconocen patrones en el lenguaje. A pesar de la aparente simplicidad de la tarea, la segmentación de oraciones es un desafío en NLP debido a la complejidad del lenguaje humano y la ambigüedad que a menudo existe en las estructuras sintácticas. La segmentación de oraciones precisa es esencial para el éxito de muchas aplicaciones de NLP y sigue siendo un área de investigación activa. [5]

Tokenization: La tokenización consiste en dividir un texto en unidades más pequeñas llamadas tokens. Los tokens pueden ser palabras, signos de puntuación, números, entre otros. La tokenización es una técnica importante ya que permite a los algoritmos procesar el texto de una manera más eficiente y precisa.

El proceso de tokenización varía según el tipo de texto y el idioma utilizado. En el idioma inglés, por ejemplo, la tokenización suele ser más sencilla, ya que las palabras están separadas por espacios. En otros idiomas, como el chino o el japonés, puede ser más difícil, ya que no hay espacios entre las palabras. En estos casos, se utilizan algoritmos especializados que pueden reconocer patrones en el texto y dividirlo en tokens. [6]

Stopwords: Las stopwords son palabras comunes que se eliminan del texto durante la etapa de preprocesamiento. Estas palabras incluyen artículos, preposiciones, conjunciones y otros términos que aparecen con frecuencia en el lenguaje, pero que no proporcionan un significado sustancial al texto. Al eliminar las stopwords, los algoritmos de NLP pueden centrarse en las palabras clave del texto y mejorar su capacidad para extraer información significativa.

Las listas de stopwords varían según el idioma y el contexto del texto, y a menudo se personalizan para cada aplicación de NLP. Las stopwords también pueden ser útiles para reducir la dimensión del espacio de características y mejorar la eficiencia computacional en tareas de procesamiento de texto. Aunque la eliminación de stopwords puede mejorar

la precisión y eficiencia de los algoritmos de NLP, es importante tener en cuenta que en algunos casos, como el análisis de sentimientos, las stopwords pueden proporcionar información valiosa sobre el tono y la actitud del autor. [6]

Lexical normalization (Lemmatization/Stemming): La normalización léxica se utiliza para reducir la variabilidad en el uso de las palabras en un texto. Esto implica la estandarización de las palabras en una forma más común o regular, lo que puede mejorar la precisión y la eficiencia en aplicaciones de NLP. La normalización léxica se utiliza a menudo en la corrección ortográfica, la clasificación de texto y la recuperación de información.

El proceso de normalización léxica puede incluir la eliminación de caracteres especiales, la corrección de errores ortográficos, la conversión de palabras en diferentes formas verbales (por ejemplo, "correr" y "corrió") a una forma base común, y la eliminación de sufijos y prefijos para reducir la variabilidad morfológica de las palabras. La normalización léxica es importante porque el lenguaje natural es altamente variable y ambiguo, y la normalización puede ayudar a mejorar la precisión y la eficiencia de los algoritmos de NLP al hacer que los textos sean más consistentes y fáciles de analizar.

La lematización y el stemming son técnicas relacionadas con la normalización léxica. Ambos métodos tienen como objetivo reducir la variabilidad de las palabras en un texto, pero difieren en su enfoque y alcance.

El stemming es una técnica que consiste en reducir una palabra a su raíz o stem, eliminando sufijos y prefijos. Esto puede ayudar a reducir la variabilidad morfológica de las palabras en un texto y mejorar la eficiencia computacional al reducir la cantidad de palabras únicas en un conjunto de datos. Sin embargo, el stemming puede no ser tan preciso como la lematización, ya que a menudo produce stems que no son palabras reales (Por ejemplo: automatización, autómatas -> automat).

Por otro lado, la lematización implica la identificación de la forma base o lemma de una palabra, que puede ser un sustantivo, un verbo, un adjetivo, entre otros. La lematización puede ser más precisa que el stemming, ya que identifica formas reales de palabras en un diccionario lema (Por ejemplo: Tengo, tienes->tener). La lematización también puede ayudar a reducir la variabilidad sintáctica de las palabras en un texto, pero puede ser más costosa en términos computacionales que el stemming. [6]

N-grams: Los n-gramas son una técnica que se utiliza para extraer información de un texto en función de los patrones de frecuencia de las secuencias de palabras. Un n-grama es una secuencia de n palabras consecutivas que se extraen de un texto, donde n es un número entero positivo. Los n-gramas se utilizan en una variedad de aplicaciones, como la generación de texto, la corrección de errores ortográficos y la clasificación de texto.

La técnica de n-gramas implica dividir un texto en todas las secuencias de n palabras consecutivas y contar cuántas veces aparece cada secuencia en el texto. Por ejemplo, si n es igual a 2, se extraen todas las secuencias de dos palabras del texto, como "procesamiento del", "del lenguaje", "lenguaje natural", etc. Los n-gramas se utilizan comúnmente para modelar el lenguaje, donde se analiza la frecuencia de las secuencias de palabras en un corpus de texto para predecir la probabilidad de ocurrencia de una secuencia en particular. Calcula la probabilidad de que una palabra siga, según las n palabras previas. [6]

$$\mathbb{P}(w_1 w_2 \dots w_n) = \prod_{i=1}^n \mathbb{P}(w_i | w_1 w_2 \dots w_{i-1})$$

Y utilizando la hipótesis de Markov, que consiste en asumir que la probabilidad de una palabra depende sólo de la palabra previa, el cálculo se puede simplificar a la siguiente expresión.

$$\mathbb{P}(w_1 w_2 \dots w_n) \approx \prod_{i=1}^n \mathbb{P}(w_i | w_{i-1})$$

Term Frequency (TF): La frecuencia de término, es una métrica utilizada para medir la importancia de una palabra en un documento. Esta métrica indica cuántas veces aparece una palabra específica en un documento en comparación con el número total de palabras en el mismo.

$$TF(t, d) = \frac{\#Veces \text{ que aparece término } t \text{ en documento } d}{\text{Total de términos del documento } d}$$

La frecuencia de término puede ser útil en muchas aplicaciones, como la recuperación de información, el análisis de sentimientos y la clasificación de texto. Sin embargo, es importante tener en cuenta que la frecuencia de término no es suficiente para medir completamente la importancia de una palabra en un documento. Es posible que una palabra aparezca con frecuencia en un documento, pero no sea relevante para el tema o contexto de este. [7]

Bag of Words: es una técnica utilizada para representar un texto como una colección de palabras independientes de su orden. El modelo BoW trata un texto como una "bolsa" de palabras y cuenta cuántas veces aparece cada palabra en el texto. Este modelo se utiliza comúnmente en tareas como la clasificación de texto, la recuperación de información y la minería de opiniones.

El modelo BoW implica dos pasos principales: tokenización y vectorización. En la tokenización, el texto se divide en palabras individuales y se eliminan las palabras vacías y de puntuación. En la vectorización, se crea un vector para cada documento o texto en función de las palabras que aparecen en él y su frecuencia. Cada palabra se convierte en una característica en el vector, y la frecuencia de cada palabra se utiliza para establecer su peso en el vector. El modelo BoW es un enfoque simple pero efectivo para el procesamiento de texto y ha demostrado ser útil en una amplia gama de aplicaciones de NLP. [7]

Inverse Document Frequency (IDF): La ponderación inversa del documento es una métrica para medir la relevancia de una palabra en un conjunto de documentos. La idea detrás del IDF es que una palabra que aparece con frecuencia en un documento, pero raramente en otros documentos, es más relevante para ese documento en particular.

$$IDF(t) = \frac{\text{Total de Documentos}}{\#Documentos \text{ donde aparece el término } t}$$

El IDF se calcula tomando el logaritmo inverso del número de documentos en el conjunto de datos que contienen la palabra específica dividido por el número total de documentos en el conjunto. La ponderación inversa del documento es útil en muchas

aplicaciones, como la clasificación de texto y la recuperación de información, porque ayuda a reducir el peso de las palabras comunes y aumentar el peso de las palabras menos comunes, pero más relevantes. La ponderación inversa del documento se utiliza a menudo en conjunto con la frecuencia de término (TF) para calcular el puntaje TF-IDF, una métrica de relevancia de palabras que es ampliamente utilizada en el NLP. [7]

Term Frequency-Inverse Document Frequency (TF-IDF): Es una métrica utilizada para medir la relevancia de una palabra en un documento. Combina la frecuencia de término (TF) de una palabra en un documento con la ponderación inversa del documento (IDF) de esa palabra en un conjunto de documentos.

$$TF - IDF(t, d) = TF(t, d) \cdot IDF(t)$$

La métrica se utiliza comúnmente en aplicaciones como la recuperación de información y la clasificación de texto. Esta métrica asigna un puntaje más alto a las palabras que son frecuentes en un documento específico pero raras en otros documentos, lo que indica que son más relevantes para ese documento. Al combinar la frecuencia de término y la ponderación inversa del documento, ayuda a mejorar la precisión y la eficacia de los algoritmos al proporcionar una medida más precisa de la relevancia de una palabra en un documento. [7]

Part of Speech Tagging (POS): Es un proceso utilizado para identificar y etiquetar las partes de la oración de cada palabra en un texto. Las partes de la oración incluyen sustantivos, verbos, adjetivos, adverbios, pronombres, preposiciones y conjunciones. El POS tagging se utiliza para ayudar a entender la estructura gramatical de un texto y proporcionar información valiosa para otras aplicaciones de NLP como la extracción de información y el análisis de sentimientos.

Este se realiza mediante el uso de algoritmos de aprendizaje automático y diccionarios de palabras que asignan una etiqueta a cada palabra en función de su función en la oración. Los algoritmos de POS tagging consideran el contexto en el que se encuentra la palabra y las palabras que la rodean para determinar su etiqueta. La precisión del POS tagging puede variar según el idioma y el contexto del texto, y puede haber ambigüedad en algunos casos en los que una palabra puede tener diferentes etiquetas dependiendo del contexto en el que se encuentra. [8]

Named Entity Recognition (NER): Se utiliza para identificar y clasificar las entidades nombradas en un texto. Las entidades nombradas son objetos o conceptos específicos que tienen un nombre propio, como personas, lugares, organizaciones, fechas y números. El NER es un paso importante en muchas aplicaciones, como la extracción de información, la recuperación de información y la minería de opiniones.

El NER se realiza utilizando algoritmos de aprendizaje automático y diccionarios de entidades previamente definidos para identificar y etiquetar las entidades nombradas en un texto. El algoritmo examina cada palabra en el texto y determina si es una entidad nombrada o no. Si se identifica una entidad nombrada, se le asigna una etiqueta que indica su tipo, como persona, lugar u organización. El NER es un proceso complejo y preciso, ya que debe tener en cuenta el contexto y las relaciones entre las palabras en un texto para identificar correctamente las entidades nombradas. [9]

Modelos de Espacio Vectorial y Embeddings de Palabras: En el ámbito de la representación algebraica de documentos de texto, existen dos modelos principales: el

Modelo de Espacio Vectorial (Vector Space Model) [10] y los Embeddings de Palabras (Word Embedding) [11]. El Modelo de Espacio Vectorial asigna una ponderación de relevancia a cada palabra en un documento. Esta ponderación puede basarse en la frecuencia de la palabra en el documento, la frecuencia de la palabra en toda la colección de documentos, o una combinación de ambas. Por otro lado, los Embeddings de Palabras involucran la representación de palabras en un espacio de baja dimensión (es decir, un vector denso de números) de manera que palabras similares en contexto estén cercanas en el espacio vectorial.

Herramientas y técnicas específicas: Skip-Gram y CBOW (Continuous Bag of Words) son dos algoritmos populares utilizados para crear embeddings de palabras. Word2Vec es una herramienta que utiliza estos algoritmos para generar representaciones vectoriales de palabras a partir de grandes cantidades de texto [11]. Doc2Vec es una extensión de Word2Vec que permite generar representaciones vectoriales no solo de palabras, sino también de frases o documentos completos [12]. El Análisis Semántico Latente (Latent Semantic Analysis, LSA) [13] y la Asignación Latente de Dirichlet (Latent Dirichlet Allocation, LDA) [14] son técnicas de reducción de dimensionalidad y modelado de temas, respectivamente, que también son fundamentales para el procesamiento y análisis de textos. Finalmente, SyntaxNet es un modelo de análisis sintáctico que identifica las relaciones gramaticales entre las palabras en una oración [15]. Todas estas herramientas y técnicas son esenciales para diversas aplicaciones de procesamiento del lenguaje natural, como la búsqueda semántica, la clasificación de documentos y el análisis de sentimientos.

2.3.2 Estado del Arte

A lo largo de los años, ha habido avances significativos en NLP, pero uno de los hitos más importantes fue la introducción de los modelos Transformers. En este estado del arte, exploraremos el panorama de NLP antes y después de los Transformers, así como el paper seminal que los introdujo.

Antes de los Transformers, los modelos de NLP se basaban principalmente en enfoques tradicionales como las cadenas de Markov ocultas (HMM), los modelos de lenguaje basados en gramáticas y los modelos estadísticos. Estos enfoques demostraron ser efectivos en tareas específicas como el etiquetado de partes del discurso, el análisis sintáctico y la traducción automática, pero tenían limitaciones en términos de escalabilidad y la capacidad de capturar dependencias a largo plazo en el lenguaje.

Además, se utilizaron modelos basados en redes neuronales, como las redes neuronales recurrentes (RNN) y las redes neuronales convolucionales (CNN), que mostraron resultados prometedores en ciertas tareas de NLP. Sin embargo, las RNNs sufrían del problema de desvanecimiento o explosión del gradiente, lo que dificultaba la captura de dependencias a largo plazo. Las CNNs, por otro lado, no eran adecuadas para modelar la secuencialidad del lenguaje.

En 2017, investigadores de Google, publicaron el paper "Attention is All You Need" [16] que introdujo el modelo Transformer, revolucionando el campo de NLP. Los Transformers presentan un enfoque innovador para el modelado del lenguaje mediante el uso de mecanismos de atención en lugar de RNNs o CNNs.

Los Transformers se basan en el concepto de "atención" para capturar las relaciones entre las palabras en un texto. A diferencia de las RNNs, que procesan la secuencia de palabras de manera secuencial, los Transformers pueden procesar varias palabras en paralelo utilizando "atención". Esto permitió capturar dependencias a largo plazo de manera más efectiva y mejorar el rendimiento en tareas como la traducción automática, el resumen de texto, la generación de texto y el análisis de sentimientos.

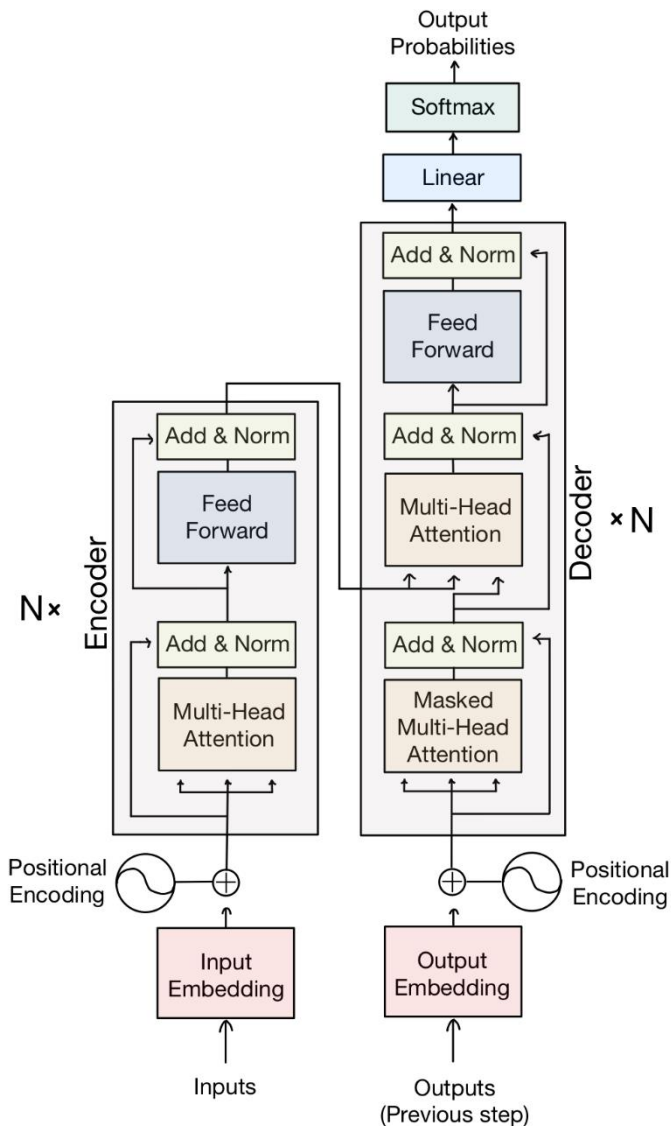


Ilustración 3: Arquitectura Transformer basado en "Attention is all you need"

La arquitectura Transformer es una innovación esencial en el procesamiento del lenguaje natural y es la piedra angular de muchos modelos de lenguaje modernos. Sus componentes son:

Input/Output Embedding: Las palabras o tokens se convierten en representaciones vectoriales densas de dimensión d_{model} , lo que permite al modelo aprender relaciones semánticas entre palabras.

Positional Encoding: Dado que el transformador no tiene una noción inherente de la posición de una palabra en una secuencia, se añade información de posición a los

embeddings. Aunque hay varias alternativas para realizarse, en esa ocasión se utilizaron las funciones seno y coseno, donde pos es la posición, i es la dimensión y d_{model} es la dimensión del embedding.

$$PE_{(pos,2i)} = \sin(pos/10000^{2i/d_{model}})$$

$$PE_{(pos,2i+1)} = \cos(pos/10000^{2i/d_{model}})$$

Encoder: Compuesto de $N = 6$ capas idénticas, cada capa contempla dos subcapas:

- Auto atención multi cabeza: Permite que el modelo preste “atención” a diferentes partes de la secuencia de entrada de manera simultánea. Matemáticamente, la atención se calcula utilizando la siguiente fórmula:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Donde Q , K y V son las matrices de consulta, llave y valor, respectivamente y d_k es la dimensión de las llaves. Para multi-cabeza, esta atención se calcula múltiples veces con diferentes pesos y luego se concatenan y proyectan:

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^O$$

$$where\ head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$$

- Red neuronal feed-forward: Consiste en dos transformaciones lineales con una activación ReLU entremedio, ayuda a transformar la representación de los datos.

$$FFN(x) = ReLU(xW_1 + b_1)W_2 + b_2$$

Cada subcapa contempla adicionalmente conexiones residuales, que ayudan a evitar el problema del desvanecimiento del gradiente en redes profundas, y una normalización que asegura que las activaciones no se vuelvan demasiado grandes o pequeñas.

Decoder: También compuesto de $N = 6$ capas idénticas, contempla adicionalmente una tercera subcapa de atención multi cabeza que se aplica al resultado del Encoder, y donde ciertas posiciones son "enmascaradas" o bloqueadas para evitar que el modelo "vea" futuros tokens en la secuencia. Es esencial para tareas como la predicción de la siguiente palabra.

Linear & Softmax: La salida del modelo se pasa por una capa lineal seguida de una función softmax para generar probabilidades para cada posible siguiente palabra en la tarea de predicción.

Una de las claves del éxito de esta arquitectura radica en el mecanismo de atención, que permite que cada palabra en la secuencia de salida asigne ponderaciones a todas las palabras en la secuencia de entrada, basándose en su relevancia. Esta ponderación dinámica facilita tareas como la traducción, donde el orden y la relación entre palabras en diferentes idiomas puede variar considerablemente.

Después de la introducción de los Transformers, surgieron varias variantes y extensiones desarrolladas por grandes empresas, para mejorar su rendimiento y aplicabilidad en múltiples tareas de NLP. Estos fueron pre entrenados mediante enormes cantidades de datos extraídos principalmente de Internet, es decir, miles de millones de palabras, lo equivalente a más de 500GB de texto, debido a estas características que requieren de un vasto consumo computacional para adquirir un conocimiento general del lenguaje, es por lo que han sido denominados *LLM (Large Language Model)*. Algunos de los más relevantes en perspectiva, fueron:

BERT (Bidirectional Encoder Representations from Transformers): En 2018, el equipo de investigación de Google presentó BERT, que propuso un modelo de lenguaje pre-entrenado bidireccional basado en Transformers. BERT revolucionó el campo de NLP al demostrar un rendimiento sobresaliente en diversas tareas. Por ejemplo, en reconocimiento de entidades nombradas BERT fue capaz de identificar nombres de personas, lugares u organizaciones en el texto con mayor precisión que los enfoques anteriores. Además, en clasificación de texto BERT superó a los modelos existentes al asignar categorías correctas a fragmentos de texto, como clasificar noticias en categorías temáticas. También en análisis de sentimientos BERT mejoró la capacidad de comprender y clasificar la polaridad de las opiniones expresadas en el texto. [17]

GPT (Generative Pre-trained Transformer): OpenAI introdujo GPT en 2018, utilizando Transformers para el modelado de lenguaje generativo. Los modelos GPT destacaron por su habilidad para generar texto coherente y natural. Esto tuvo un impacto significativo en áreas como la redacción automática, donde GPT fue capaz de generar artículos y ensayos completos que sonaban escritos por humanos. En el ámbito del diálogo generativo, GPT demostró una capacidad impresionante para mantener conversaciones fluidas y responder de manera relevante a preguntas y solicitudes. Además, en la creación de historias, GPT se destacó por generar narrativas convincentes y cautivadoras. [18]

GPT-3: En 2020, OpenAI lanzó GPT-3, uno de los modelos de lenguaje más grandes y potentes hasta la fecha. Con 175 mil millones de parámetros, GPT-3 alcanzó resultados sorprendentes en diversas tareas de NLP. Por ejemplo, en traducción automática, GPT-3 fue capaz de traducir textos entre diferentes idiomas con un alto grado de precisión y fluidez. En generación de código, GPT-3 demostró la capacidad de escribir fragmentos de código funcional en diferentes lenguajes de programación. Además, en la creación de poemas, GPT-3 sorprendió con su habilidad para generar versos líricos que siguieron estructuras y estilos poéticos reconocibles. [19]

T5 (Text-to-Text Transfer Transformer): En 2020, Google AI presentó T5, un modelo que adopta un enfoque "text-to-text" y puede abordar múltiples tareas de procesamiento del lenguaje natural utilizando una sola arquitectura y un solo entrenamiento. T5 se entrenó en una escala masiva y ha demostrado un rendimiento sobresaliente en tareas de generación de lenguaje, traducción automática, extracción de información, preguntas y respuestas, entre otras. [20]

La introducción de los Transformers ha tenido un impacto significativo en el campo del Procesamiento del Lenguaje Natural. Estos modelos han permitido avances en tareas de NLP que antes eran desafiantes o incluso inalcanzables. Desde el paper seminal "Attention is All You Need" hasta las variantes y extensiones posteriores, los Transformers han demostrado ser una poderosa herramienta para el modelado del lenguaje. Con su capacidad para capturar relaciones a largo plazo y su aplicabilidad en una amplia gama de

tareas, los Transformers han llevado a un nuevo nivel de comprensión y generación de lenguaje humano.

Desde entonces, en base a esta nueva arquitectura se han lanzado diversos chatbots muy potentes con los que se puede interactuar desde cualquier navegador, donde a la fecha, destacan los siguientes:

ChatGPT4 de OpenAI [21] es actualmente considerado el modelo líder en el campo, sin ningún modelo que se le compare. Aunque su conocimiento se limita al año 2021, posee la impresionante cantidad de 1.7 billones de parámetros y una ventana de contexto de hasta 8 196 tokens, con estas características destaca por su capacidad de respuesta en una amplia variedad de temas a nivel global, sobresaliendo en calidad y extensión de respuesta. Este modelo abarca prácticamente todas las funciones de modelos anteriores de NLP y va más allá. Sus características destacadas incluyen la traducción y corrección de idiomas, la capacidad de explicar temas complejos con detalle y personalización, la generación de resúmenes, la habilidad creativa para escribir canciones, poemas y artículos científicos, así como el razonamiento y la resolución de problemas, entre otras. Actualmente sólo es accesible mediante una suscripción de pago mensual, y limita su capacidad a 50 mensajes cada 3 horas.

Bing Chat de Microsoft, integrado exclusivamente con su navegador, Microsoft Edge, utilizando tecnología brindada por OpenAI, es capaz de brindar respuestas en tiempo real ya que busca en internet para mejorar sus respuestas, actualmente limita las conversaciones a 30 mensajes a la vez, con una capacidad máxima de escritura de apenas 4 000 caracteres (~800 palabras).

Bard de Google AI, con sus 136 mil millones de parámetros, es capaz de brindar respuestas en tiempo real ya que tiene acceso a internet para tratar de mejorar sus respuestas, actualmente tiene una ventana de contexto de apenas 1024 tokens, y es el modelo más propenso a tener “alucinaciones”, es decir, inventar información.

Ahora mencionando algunos menos conocidos están:

Claude de Anthropic, aun teniendo su acceso mayoritariamente restringido a unos pocos y de pago, se le considera uno de los modelos más cercanos a rivalizar con ChatGPT con su nivel de respuesta.

LLaMa de Meta AI [22], lanzó modelos que van desde 7 a 65 mil millones de parámetros entrenados exclusivamente con datos públicos, esto sentó las bases y dio espacio para acelerar el desarrollo de modelos LLM Open Source más económicos, tales como Alpaca de la Universidad de Stanford [23], Vicuna de LMSYS Org [24], entre otros, con costos de entrenamiento menores a mil dólares.

Los avances actuales de los LLM son asombrosos; han logrado un rendimiento sin precedentes en diversas tareas, como el modelado de lenguaje, la traducción automática, el análisis de sentimientos y la clasificación, entre otros. Estos modelos son capaces de entender y generar texto coherente en varios idiomas y procesar grandes volúmenes de información en poco tiempo. Al entrenarse con conjuntos de datos vastos, pueden capturar patrones complejos y proporcionar respuestas bien fundamentadas. Además, los LLM han mostrado habilidades notables en la generación de texto creativo, como escribir poesía, cuentos e incluso código de programación.

No obstante, los LLM también tienen limitaciones. Son modelos de gran escala que requieren recursos computacionales considerables y conjuntos de datos de entrenamiento extensos. El entrenamiento y la implementación de estos modelos pueden ser costosos en términos de tiempo y potencia de cálculo, limitando así su accesibilidad para ciertos entornos o aplicaciones con recursos restringidos.

A pesar de su habilidad para generar texto coherente, aún pueden cometer errores y generar información incorrecta o no verificada. También son susceptibles a los sesgos presentes en los datos con los que fueron entrenados, lo que puede resultar en respuestas parciales o discriminatorias. Debido a su arquitectura compleja y la falta de transparencia en las decisiones de atención, pueden ser difíciles de interpretar, lo que plantea desafíos en la explicabilidad y confiabilidad de los resultados generados por estos modelos.

En este contexto, OpenAI se distinguió de sus competidores al implementar el "Reinforcement Human Learning Feedback" (RHLF), un enfoque que combina aprendizaje automático y retroalimentación humana para mejorar la capacidad de los modelos de lenguaje. En este proceso, los modelos se entrenan inicialmente con grandes cantidades de datos y luego se afinan con retroalimentación humana para corregir errores, reducir sesgos y mejorar la coherencia y relevancia de las respuestas generadas. Este método es esencial para mejorar la precisión y fiabilidad de los modelos de lenguaje, ya que permite ajustar y corregir errores y deficiencias que no pueden ser detectados o corregidos únicamente con métodos automatizados. Aunque los LLM son extremadamente poderosos y capaces, todavía pueden generar respuestas incorrectas, incoherentes o sesgadas. La retroalimentación humana es crucial para asegurar que los modelos estén alineados con nuestros valores, normas y expectativas, y para garantizar que sean útiles, confiables y justos en una amplia variedad de aplicaciones.

A pesar de que los LLM pueden procesar y generar texto en lenguaje natural, no tienen una comprensión profunda de los conceptos ni poseen conocimiento del mundo más allá de lo aprendido durante el entrenamiento.

En términos de interacción con los usuarios, los LLM también presentan limitaciones. Pueden tener dificultades para manejar preguntas o instrucciones ambiguas y, en ocasiones, pueden dar respuestas evasivas o irrelevantes. Además, pueden carecer de la capacidad de realizar un seguimiento coherente de una conversación a largo plazo y pueden olvidar información previa en un intercambio.

Es importante tener en cuenta estas limitaciones al utilizar los LLM y siempre es recomendable verificar la información que proporcionan y considerar sus respuestas como sugerencias o puntos de partida, en lugar de respuestas definitivas.

Cabe mencionar adicionalmente, que particularmente durante este año, han ocurrido una significativa cantidad de avances con respecto de estos modelos, en comparación a años anteriores, elevando rápidamente las expectativas del estado del arte dentro del corto plazo.

Dentro del futuro de NLP de código abierto, es importante mencionar las siguientes organizaciones:

La empresa Hugging Face [25], fundada en 2016 con el objetivo de facilitar el acceso a la inteligencia artificial y a los modelos de procesamiento del lenguaje natural (NLP). Hugging Face ha desarrollado una plataforma y una biblioteca de software llamada

"Transformers" que se ha vuelto muy popular en la comunidad de NLP. También proporciona servicios y herramientas relacionadas con el desarrollo de aplicaciones de NLP, incluyendo su hub de modelos, donde los usuarios pueden compartir y acceder a modelos preentrenados, y su API llamada "Hugging Face Inference API", que permite a los desarrolladores implementar y escalar modelos de NLP en producción. Hugging Face ha obtenido reconocimiento y ha recibido inversiones por parte de importantes empresas e inversores en el campo de la inteligencia artificial. Su enfoque en la comunidad y en la democratización del acceso a los modelos de NLP ha contribuido a su crecimiento y éxito en el ámbito empresarial.

La organización Large Model Systems Organization (LMSYS Org) [26] es una organización de investigación abierta fundada por estudiantes y profesores de UC Berkeley en colaboración con UCSD y CMU. Tiene por objetivo hacer que los LLM sean accesibles para todos mediante el desarrollo conjunto de modelos abiertos, conjuntos de datos, sistemas y herramientas de evaluación. A la fecha se adjudica el desarrollo del modelo Vicuña, basado en el modelo LLaMA de Meta AI, y la plataforma Chatbot Arena [27], que permite probar y comparar los diversos modelos OpenSource más populares del momento.

2.3.3 Legal Tech

Legal Tech es un término que hace referencia a la aplicación de tecnología en el ámbito jurídico, englobando una variedad de herramientas y tecnologías, desde la automatización de procesos hasta la inteligencia artificial y el procesamiento del lenguaje natural (NLP, por sus siglas en inglés), siendo esta última una de las herramientas más empleadas. El propósito principal de Legal Tech es hacer que el derecho sea más accesible, eficiente y efectivo. [28]

Las herramientas basadas en NLP facilitan a los profesionales jurídicos la automatización de tareas que antes consumían mucho tiempo, como la revisión de documentos y contratos. Además, pueden identificar términos específicos y patrones recurrentes en textos legales, lo que facilita a los abogados la identificación rápida de posibles problemas y riesgos.

Más allá de la revisión de documentos, el NLP también puede ser usado para análisis de sentimientos e identificación de tendencias en la comunicación. Por ejemplo, las compañías de seguros pueden emplear tecnologías de NLP para analizar las solicitudes de indemnización de los clientes y determinar si existen indicios de fraude.

Asimismo, el NLP puede ser útil en la resolución de disputas y arbitraje. Los sistemas basados en NLP pueden analizar grandes volúmenes de datos para identificar tendencias en los casos, lo que puede ayudar a abogados y jueces a tomar decisiones más informadas y justas.

En resumen, la combinación de Legal Tech y técnicas de NLP tiene el potencial de transformar la práctica del derecho. Al permitir una mayor automatización de tareas y una mayor eficiencia en la toma de decisiones, estas herramientas pueden mejorar el acceso a la justicia y hacer que los servicios jurídicos sean más accesibles y asequibles para todos.

El modelo BERT, ha sido notorio para el desarrollo de esta área, desde su salida, se han desarrollado variaciones de este modelo llamados LEGAL-BERT [29], los cuales han sido entrenados con grandes cantidades de documentos legales a lo largo del mundo, mayoritariamente en países de habla inglesa, sin embargo, también se han reportado variaciones desarrolladas para Brasil [30], Italia [31], India [32], Israel [33], entre otros.

La última gran noticia del área fue la salida de dos nuevos asistentes legales privados que utilizan inteligencia artificial:

- CoCounsel, de CaseText (Estados Unidos) [34]: La empresa afirma que este modelo creado usando el modelo GPT-4 de OpenAI, que fue personalizado para la industria legal, es capaz de generar interpretaciones avanzadas de documentos legales completos, siendo también capaz de referenciarlos, ayuda a preparar declaraciones identificando tópicos relevantes y preguntas para cada uno, es capaz de buscar información particular entre los diversos documentos, e incluso puede resumir documentos sin perder los puntos claves.
- Sofía, de Tirant (Chile) [35]: La empresa española afirma que este modelo es capaz de analizar cualquier tipo de documento y expediente, de realizar análisis predictivos y de recomendación, siendo capaz de ofrecer la legislación, formularios, jurisprudencia y doctrina relacionada gracias a una búsqueda conceptual inteligente.

Ahora bien, estos avances han sido mayoritariamente en el extranjero o por privados, y poco o nulos avances existen hoy en nuestro país, la iniciativa pública en este ámbito parece estancada. Estos desarrollos privados son testigos de la capacidad de los algoritmos para desentrañar el lenguaje jurídico, pero la falta de alternativas o desarrollo de carácter público amplifica la relevancia de este estudio, es por esto que este proyecto busca ser el punto de partida para futuros avances nacionales.

2.3.4 Herramientas para el proyecto

Hasta la fecha, se pudo apreciar que ChatGPT, basado en la arquitectura GPT-3.5 desarrollada por OpenAI, tiene la capacidad de generar textos ilustrativos y concisos en el estilo de las sentencias emitidas por los tribunales chilenos, con mayor parecido que cualquier otro modelo disponible, y si se le entrega demandas y contestaciones relativamente cortas, es capaz de redactar una sentencia coherente, sin embargo, es importante destacar que este modelo no fue específicamente entrenado para tal propósito y sus habilidades en este ámbito son limitadas.

En lo que respecta a la capacidad de procesamiento de texto, se ha establecido que ChatGPT 3.5 puede interpretar hasta un máximo de 4096 tokens simultáneamente, lo cual es impresionante pero también implica ciertas restricciones. Dado que un caso judicial típico puede abarcar una cantidad considerable de información, esta capacidad podría resultar insuficiente para analizar y comprender plenamente el contexto de un caso complejo.

Es importante destacar que OpenAI no ha actualizado la base de entrenamiento del modelo más allá del año 2021, por lo que su conocimiento se limita a la información disponible hasta dicha fecha.

Si bien el desarrollo de modelos de lenguaje como ChatGPT muestra prometedoras perspectivas para el futuro y muestra los alcances de las posibilidades técnicas, es importante reconocer que replicar modelos similares en una computadora personal resultaría extremadamente desafiante debido a las enormes demandas de recursos computacionales e infraestructurales que se requieren para entrenar y ajustar modelos de lenguaje a gran escala.

En la actualidad, las mejores alternativas Open Source vienen en distintos tamaños para correr de manera local, los mejores modelos a la fecha son: LLaMA-2 70B [36] y Falcon 40B [37], sin embargo, estos exigen una memoria RAM de entre 85 y 100 GB para su manejo satisfactorio, por otro lado, los mejores modelos de tamaño medio son: Vicuna-13B y WizardLM-13B, que exigen una memoria RAM de más de 16 GB para su manejo satisfactorio, quedando entonces para los computadores de uso común, sólo como opción los modelos más pequeños, de 7 mil millones de parámetros, pero estos tienen una potencia muchísimo más reducida, quedando en evidencia su menor capacidad de respuestas, al compararse con los 175 mil millones de parámetros que contempla el modelo de ChatGPT.

Actualmente, existen plataformas en línea que ofrecen la posibilidad de probar gratuitamente las capacidades de varios modelos de código abierto en tiempo real, como el chat de LMSYS (Large Model Systems Organization) y el chat h2oGPT [38], mantenido por H2O.ai, una compañía de software estadounidense. Fue a través de estas plataformas que se logró comparar los resultados de los modelos, revelando que ChatGPT es el único capaz de redactar una sentencia típica de los tribunales chilenos. Además, este es uno de los pocos modelos capaces de inferir significado incluso entre palabras mal escritas. Por lo tanto, se seleccionó este modelo como base para explorar las hipótesis del proyecto.

Existen dos métodos principales para potenciar las capacidades de un LLM.

Ajuste Fino (Finetuning): Consiste en tomar un modelo preentrenado y continuar su entrenamiento con un conjunto de datos específico o en un dominio particular. La premisa de este enfoque es que un modelo que ya ha aprendido patrones y relaciones generales del lenguaje puede ser adaptado para comprender mejor y responder de manera más adecuada a preguntas o tareas en un ámbito específico. Por ejemplo, un LLM general puede ser afinado con textos médicos para mejorar su capacidad de responder preguntas médicas. El ajuste fino permite personalizar el modelo para adaptarlo mejor a las necesidades de una aplicación específica, mejorando así la precisión y relevancia de las respuestas generadas.

Incorporación de Bases de Conocimiento (Knowledge Base Embedding): Implica integrar información estructurada, como la que se encuentra en bases de conocimiento, en el modelo de lenguaje. Las bases de conocimiento contienen información organizada y estructurada sobre el mundo, y su integración en un LLM puede ayudar a mejorar la calidad y precisión de las respuestas generadas por el modelo. Esto se debe a que el modelo no solo se basa en los patrones lingüísticos aprendidos durante su entrenamiento inicial, sino que también tiene acceso a información estructurada y verificada. El proceso de incorporación de bases de conocimiento implica técnicas

avanzadas de procesamiento de lenguaje natural y puede ser bastante complejo, pero es una forma efectiva de potenciar la capacidad de un LLM para responder preguntas de manera precisa y completa.

La principal diferencia práctica entre estos métodos radica en que el primero se utiliza para imitar un estilo de comportamiento o respuesta, mientras que el segundo sirve para adquirir conocimiento sobre hechos o conceptos. Para este proyecto en particular, se utiliza el ajuste fino con el objetivo de imitar de manera más elaborada o estructurada, el tipo de razonamiento o respuesta de la Corte Suprema de Chile.

Capítulo 3

Levantamiento de Requerimientos

En este capítulo, exploraremos las fases del proyecto, detallando sus iteraciones y examinando las causas específicas de su descarte. Este proceso de evaluación crítica fue fundamental para la identificación precisa de los requerimientos funcionales del proyecto, asegurando una alineación óptima con los objetivos establecidos. A continuación, se presenta un desglose de dichas iteraciones, proporcionando una visión clara de cómo cada paso contribuyó al refinamiento de los requerimientos y al éxito eventual de la iniciativa.

1. Análisis de factores o cláusulas comunes en contratos llevados a juicio

Inicialmente, se procuró compilar un amplio corpus de contratos que habían sido objeto de litigios para correlacionarlos con las sentencias judiciales pertinentes. El propósito era identificar patrones en las cláusulas que frecuentemente condujeran a malentendidos y disputas legales.

No obstante, se enfrentaron obstáculos significativos debido a la naturaleza privada de los contratos, lo que dificultaba su recopilación sistemática. Además, tras intercambios esclarecedores con los abogados involucrados en el proyecto, se reveló que este enfoque inicial era inadecuado por varias razones:

- Un solo contrato podría desencadenar múltiples demandas, cada una basada en interpretaciones distintas de sus cláusulas, lo cual complica la correlación directa con resultados legales específicos.
- A menudo, las disputas se centran en segmentos muy particulares del documento, lo que disminuye el valor de analizar el contrato en su totalidad.
- No siempre existe una relación directa y evidente entre la sentencia judicial y las cláusulas específicas del contrato en cuestión.
- Este método no toma en cuenta los argumentos presentados por cada parte involucrada en el litigio, los cuales son cruciales para comprender el contexto completo de cada caso.

A raíz de estos hallazgos, fue necesario replantear la estrategia de investigación para abordar estas complejidades de manera más efectiva. Se tomó la decisión de estudiar las demandas, contestaciones y sentencias asociados al contrato como tal.

2. Entrenamiento de un modelo utilizando demandas, contestaciones y sentencias de la Corte Suprema

Se les encargó a los expertos del área legal recopilar el nuevo corpus, y lograron brindar 10 casos civiles en total. Estos casos estaban en formato PDF, y varios de éstos eran en realidad documentos escaneados. De los casos y sus documentos se apreciaron los siguientes aspectos:

Caso	Categoría	Demanda	Contestación	Sentencia
Primero	Restitución de garantía de arrendamiento	6 páginas	4 páginas	6 páginas
Segundo	Resolución de contrato con indemnización de perjuicios	40 de las 70 páginas, contienen evidencia adicional ilegible o páginas en blanco	9 páginas	33 páginas
Tercer	Cobro de pesos en juicio sumario	3 páginas	<u>Audiencia de contestación con poca o nula información.</u> 1 página	5 páginas
Cuarto	Incumplimiento de contrato	Más de la mitad de las 50 páginas es evidencia adicional	29 páginas	40 páginas
Quinto	Cumplimiento de contrato con indemnización de perjuicios	29 páginas	7 páginas	83 páginas
Sexto	Cobro de pagaré	<u>No se encuentra</u>	15 páginas	<u>No se encuentra</u>
Séptimo	Cobro de honorarios	6 páginas	5 páginas	7 páginas
Octavo	Incumplimiento de contrato e indemnización	12 de las 32 páginas es evidencia adicional	15 páginas	<u>No se encuentra</u>
Noveno	Cobro de honorarios	7 páginas	<u>24 páginas ilegibles</u>	<u>No se encuentra</u>
Décimo	Cobro de honorarios	7 páginas	<u>Audiencia de contestación con poca o nula información.</u> 1 página	8 páginas

Tabla 1: Casos Judiciales Adquiridos

En base a estas apreciaciones, se decide tratar con aquellos casos que poseen la información más completa, por lo que se procede a buscar plataformas que permitan extraer el contenido textual de estos documentos PDF.

Para rescatar el contenido textual de documentos escaneados, fue necesario probar distintas herramientas, de la cual terminó destacando ilovepdf [39] por su proceso de OCR (Optical Character Recognition) que permite identificar automáticamente símbolos o caracteres a partir de imágenes, para luego interactuar con ellos mediante un programa de edición de texto o similar, en este caso, Word, para luego extraer ese contenido en Python usando la librería docx2txt.

Se notó que la extracción no logró ser perfecta, y los documentos presentaron una cantidad notoria de información perdida o mal interpretada, y en ocasiones ilegibles.

En este punto se identificaron los siguientes obstáculos con seguir optando por esta alternativa:

- De los modelos más avanzados que se encontraron, sólo pueden interpretar hasta 4 096 tokens (~ 8 000 palabras) a la vez, lo que es muy acotado, especialmente considerando que cada uno de estos documentos por separado posee por lo menos 10 000 palabras.
- ChatGPT es el único modelo capaz de inferir el significado de algunas palabras mal escritas de acuerdo con el contexto en el que se encuentra.
- Este enfoque no considera las discusiones durante el juicio, omitiendo información valiosa y necesaria para correlacionar con la sentencia.

A raíz de estos hallazgos, se tomó la decisión de estudiar las capacidades de interpretación y de decisión de los modelos a partir de casos resumidos, creados por los expertos legales.

3. Utilización de un modelo Open Source

En el intento de conseguir una alternativa gratuita que pueda funcionar sin internet, de manera de que los datos no sean compartidos de ninguna forma a externos, es que se probaron diversas herramientas para ello, tales como freedomgpt [40] o oobabooga [41], que son plataformas que permiten utilizar modelos open source de manera local.

Sin embargo, los modelos de código abierto disponibles actualmente y que pueden ejecutarse en una computadora local estándar, no cumplen con las expectativas establecidas por los modelos avanzados de OpenAI, por las siguientes razones:

- Poco o nulo dominio de aspectos legales.
- Tiempos extensos en la generación de respuestas.
- Respuestas demasiado cortas.
- Acotada ventana de contexto.

Los modelos de OpenAI, a diferencia de los demás, presentó las siguientes capacidades empíricas, resaltando la vastedad de la base de datos utilizada para su entrenamiento y su capacidad de respuesta.

- **Dominio amplio del lenguaje**, capaz de comprender y redactar textos en formato legal, capaz de entender el significado de las palabras incluso con errores ortográficos, o con palabras escritas en código (números en lugar de vocales, palabras invertidas, etc.).
- **Conocimiento extenso de Chile**, capaz de entender y escribir en dialecto informal típico, conoce algunas leyes, entiende y es capaz de replicar la estructura y vocabulario de sentencias de tribunales chilenos.
- **Extensa ventana de contexto**, de 4 096 tokens.
- **Variante gratuita online**, de fácil acceso y uso ilimitado.

A raíz de estos hallazgos, se tomó la decisión de estudiar las capacidades de los modelos de pago, en particular los ofrecidos por OpenAI.

4. Utilización del modelo davinci OpenAI

La característica más valiosa de los modelos de OpenAI para el proyecto, es su conocimiento previo sobre el sistema judicial chileno, así como su capacidad para responder de manera adecuada. Sin embargo, tras realizar pruebas y evaluar el modelo junto con el abogado y profesor Rodrigo Coloma, se identificó que aún falta cierto grado de contexto y ejemplos para optimizar las respuestas. Al comparar el modelo predeterminado GPT-3.5 con el modelo GPT-4, se encontró que son bastante similares, aunque, al elegir entre las mejores respuestas, las del último siempre fueron preferidas.

Dado que los costos de OpenAI son relativamente accesibles para entrenamientos de menor escala, se decidió optar por esta alternativa para obtener los mejores resultados técnicos posibles con un presupuesto limitado. A la fecha de realización de esta alternativa, el modelo davinci era el mejor modelo de OpenAI disponible para el finetuning.

Sin embargo, este modelo estaba diseñado para continuar el texto previo y no para recrear una experiencia del tipo chatbot, adicionalmente el modelo tenía presupuestado que se le brindara una vasta cantidad de ejemplos con los que no se contaba, por lo que los resultados conseguidos con esta alternativa fueron muy lejanos a los esperados, con texto incoherente y vago.

Al poco tiempo de realizada esta iteración, OpenAI liberó su modelo GPT-3.5-turbo para el finetuning con una nueva metodología de entrenamiento, con la que se vislumbró la posibilidad de conseguir mejores resultados.

A raíz de estos hallazgos, se tomó la decisión de entrenar ese modelo como última alternativa. En esta ocasión, con los requerimientos funcionales mucho más claros:

Se espera que el modelo final, sea una mejora con respecto a los resultados por defecto que se pueden conseguir a la fecha, de la versión gratuita del ChatGPT de OpenAI (GPT-3.5), es decir, que mantenga o mejore las características considerando los siguientes aspectos:

- **Facilidad de acceso:** Accesible desde cualquier equipo con acceso a internet.
- **Tiempo de respuesta:** Resultados obtenidos en pocos segundos
- **Calidad de respuesta:** Resultados que tengan la extensión, calidad de redacción y utilidad definidos en los ejemplos.

Capítulo 4

Diseño y Construcción de la Solución

En este capítulo, se detalla las etapas claves del desarrollo del modelo final del proyecto de investigación. En primer lugar, se cubre el cómo se escogieron las mejores preguntas y respuestas para comenzar con el ajuste fino del modelo, en segundo lugar, se cubre el cómo se deben formatear los pares para ser introducidos en la interfaz de programación de aplicaciones (API), en tercer lugar, se cubren los comandos necesarios para iniciar el trabajo de ajuste fino, por último, se comparan y evalúan las respuestas generadas por el modelo sin entrenar y el modelo entrenado.

4.1 Comprensión de los datos

Teniendo en consideración, las capacidades actuales de los modelos de OpenAI en torno al área legal de interés, procedía revisar cómo sacarle el máximo rendimiento o uso en torno al problema inicial, el de resolver o dar lineamientos sobre cómo resolver los casos judiciales.

Para esto fue necesaria la ayuda del equipo de abogados para que seleccionaran y redactaran consultas de interés, en base a los casos más apropiados que pudieran identificar, en base a esto, se procedería a realizar una serie de comparaciones e iteraciones para tratar de ir mejorando la calidad de respuesta que se pudiera recibir del modelo.

Para la comparación y selección final de las mejores consultas y respuestas, se utilizó el siguiente procedimiento:

Se seleccionaron los dos casos de mayor interés para las consultas, de estos se crearon tres consultas en total, la primera consulta consiste en el primer caso sin contexto adicional, la segunda consulta consiste en el primer caso, pero agregando un ligero contexto adicional para comparar la diferencia de respuestas, por último, la consulta tres consiste en el segundo caso con todo un contexto previo.

Redactadas las tres consultas, se procede a comparar las respuestas de los modelos ejecutándose tres veces cada consulta por cada modelo, es decir, GPT-3.5 y GPT-4. De esta manera podría revisarse la consistencia en las respuestas por cada modelo y cómo se diferencia con el otro.

Luego se construye un documento que reúna las respuestas de cada consulta, y no se especifica a qué modelo se atribuyen, esto con el afán de al enviárselo a los abogados, estos

escojan sin ningún sesgo cuales son las mejores respuestas. A continuación, se explicitan las consultas.

- Primer caso [Sin Contexto]

La empresa de transportes “Los Palotes Limitada” se adjudicó una licitación de la empresa “Cimientos Industriales José Manuel”. En este contexto, las partes con fecha 10 de enero del año 2014 celebraron un contrato titulado: “Contrato N° 1438 de movimiento de materiales y servicio de cargadores frontales”.

Luego, en virtud de dicho contrato, la empresa Los Palotes Limitada se comprometió a prestar servicios de cargador frontal y movimiento de materiales en la planta de áridos de la comuna de Caldera, durante el plazo de cinco años, computados desde el 1 de febrero del año 2014. Lo anterior, incluyó una cláusula de prórroga y renovación automática de un año, no obstante, la facultad ambas partes de resciliar el contrato, por medio de carta certificada al domicilio de la contraparte, con a lo menos 30 días de anticipación a la fecha de término de la vigencia original del contrato o de alguna de sus prórrogas.

Sin perjuicio de lo anterior, se pactó que Cimientos Industriales José Manuel podrá rescindir anticipadamente el Contrato en cualquier momento con un preaviso mínimo de 30 días. Asimismo, que dicha facultad, no da derecho a Los Palotes Limitada a reclamar indemnización o pagos adicionales de cualquier tipo. También, se acordó que Cimientos Industriales José Manuel podrá resolver anticipadamente el contrato de pleno derecho, sin necesidad de acciones judiciales ni de medidas cautelares y sin derecho a indemnización alguna, ello siempre y cuando Los Palotes Limitada incumpliera el contrato convenido.

La retribución pactada por los servicios de Los Palotes Limitada consistió en un pago fijo mensual de \$23.760.340 para los proyectos de extracción y transporte de materiales, así como un monto variable en función del desempeño de la producción y planificación con ajustes semestrales.

Posteriormente, se acreditó, que la empresa Los Palotes Limitada incurrió en varios gastos con motivo de la ejecución del contrato, los cuales pretendía compensarlos con lo que recibiere de la demandada, obteniendo, además, un margen de ganancia. Asimismo, también consta que en las bases de la licitación se encontraba expresa la cláusula señalada, respecto a la duración del contrato.

Con fecha 2 de septiembre del año 2014, Cimientos Industriales José Manuel, mediante carta certificada protocolizada ante Notario, suscrita por su gerente general y gerente corporativo, manifestó y notificó a Los Palotes Limitada la decisión de rescindir el Contrato N° 1438 de movimiento de materiales, por que terminaría el día 2 de octubre del año 2014.

Consecuencialmente, Los Palotes Limitada deduce una demanda civil por responsabilidad contractual, en procedimiento ordinario, sobre cumplimiento de contrato con indemnización de perjuicio, en contra de Cimientos Industriales José Manuel, en virtud del contrato celebrado por las partes. Lo anterior, solicitando el cumplimiento del contrato y el pago de la suma de \$1.919.325.062, respecto de los perjuicios sufridos, más reajustes e intereses, con expresa condenación de costas.

Al respecto, Los Palotes Limitada esgrime que incurrió en costos e inversiones significativas al comenzar a prestar servicios el día 1 de febrero del año 2014. Asimismo, que con fecha 11 de septiembre del mismo año, Cimientos Industriales José Manuel le comunicó su decisión de rescindir el contrato a partir del día 2 de octubre del año 2013. Ello, mediante un ejercicio de una facultad, si bien contenida en el contrato, pero sin motivo alguno, dejando a la empresa en desprotección y sin acciones judiciales para solicitar indemnización alguna.

En este sentido, la empresa Los Palotes Limitada sostiene que la cláusula que contiene dicha facultad es anacrónica, carece de aplicabilidad y debe considerarse no escrita. Al respecto, aquella debe interpretarse teniendo en cuenta la intención de las partes, en particular en cuanto a la naturaleza del servicio, ya que, esta es la única forma de responder a la economía de contratos centrándose en la inversión, los costos y los rendimientos de los contratos. Luego, en este caso, el uso de la facultad señalada actúa como una vulneración a la buena fe contractual, la naturaleza y propósito práctico del contrato, alterando el equilibrio y la reciprocidad del contrato. Lo anterior, esgrime frustraría el fin del contrato, el propósito ideal que ambas partes tenían al entablar una relación a largo plazo, generando a Los Palotes Limitada perjuicios a título de daño emergente, lucro cesante y daño moral.

En otro término, la demandada Cimientos Industriales José Manuel argumentó que la cláusula antes señalada es válida, ya que, el contrato se adjudicó en un proceso de licitación, en el cual constaba dicha cláusula en sus bases. Asimismo, hacen referencia a que los demandantes realizaron su oferta, sin pronunciarse sobre la cláusula mencionada.

En los sucesivos, señala que no existe ambigüedad o contradicción en los términos de la cláusula. Además, agregó que la empresa Los Palotes Limitada ya había participado en otras licitaciones presentadas por Cimientos Industriales José Manuel y celebrado otros contratos que contenían las mismas o similares cláusulas.

Por último, esgrime haber ejercido la facultad del modo previsto en el contrato, asimismo, que no estaba estipulado en el contrato explicar los motivos de su decisión, pero señala que la restricción administrativa del acceso a la ribera del río de Cachapoal le impidió pagar el precio originalmente pactado. Agregó que se ofreció a modificar los términos del contrato, pero Los Palotes Limitada se negó. En otro término, acusa a la empresa de vulnerar la buena fe contractual al intentar negar el valor de la cláusula señalada, vulnerando el principio pacta sunt servanda y su pleno efecto vinculante para las partes. Finalmente, alegó la improcedencia de los perjuicios y el quantum demandado.

En consideración a los antecedentes expuestos, es que se requiere que la Corte Suprema de Chile resuelva la controversia suscitada por las partes.

- Contexto previo agregado para el primer caso

Ten en consideración los siguientes artículos del Código Civil Chileno para responder mi próxima pregunta, pero sin mencionarlos

Art. 1545. Todo contrato legalmente celebrado es una ley para los contratantes, y no puede ser invalidado sino por su consentimiento mutuo o por causas legales.

Art. 1546. Los contratos deben ejecutarse de buena fe, y por consiguiente obligan no sólo a lo que en ellos se expresa, sino a todas las cosas que emanan precisamente de la naturaleza de la obligación, o que por la ley o la costumbre pertenecen a ella.

Art. 1560. Conocida claramente la intención de los contratantes, debe estarse a ella más que a lo literal de las palabras.

Art. 1561. Por generales que sean los términos de un contrato, sólo se aplicarán a la materia sobre que se ha contratado.

Art. 1562. El sentido en que una cláusula puede producir algún efecto, deberá preferirse a aquel en que no sea capaz de producir efecto alguno.

Art. 1563. En aquellos casos en que no apareciere voluntad contraria deberá estarse a la interpretación que mejor cuadre con la naturaleza del contrato. Las cláusulas de uso común se presumen aunque no se expresen.

Art. 1564. Las cláusulas de un contrato se interpretarán unas por otras, dándose a cada una el sentido que mejor convenga al contrato en su totalidad. Podrán también interpretarse por las de otro contrato entre las mismas partes y sobre la misma materia. O por la aplicación práctica que hayan hecho de ellas ambas partes, o una de las partes con aprobación de la otra.

Art. 1565. Cuando en un contrato se ha expresado un caso para explicar la obligación, no se entenderá por sólo eso haberse querido restringir la convención a ese caso, excluyendo los otros a que naturalmente se extienda.

Art. 1566. No pudiendo aplicarse ninguna de las reglas precedentes de interpretación, se interpretarán las cláusulas ambiguas a favor del deudor. Pero las cláusulas ambiguas que hayan sido extendidas o dictadas por una de las partes, sea acreedora o deudora, se interpretarán contra ella, siempre que la ambigüedad provenga de la falta de una explicación que haya debido darse por ella.

- Segundo caso [Con Diversos Contextos]

Hola Chat GPT tengo que hacerte algunas preguntas, pero antes necesito que aprendas algunas cosas acerca de cómo decide la Corte Suprema en distintos casos:

Caso 1. GABRIEL contrata al abogado ENRIQUE. La cláusula 2ª del pacto de honorarios que suscriben ambos dice: “Los honorarios del abogado ENRIQUE, serán de un 20% de lo que obtenga en el juicio laboral encomendado. Es decir, un veinte por ciento del total de la suma de dinero que el juez declare en favor del trabajador”. En su cláusula 5ª, agrega que “En caso que GABRIEL llegue a un acuerdo de carácter judicial o extrajudicial con la demandada, el abogado ENRIQUE recibirá por concepto de honorarios un veinte por ciento, de lo que el trabajador acuerde en dicho pacto”. ENRIQUE interpone una demanda contra el antiguo empleador de GABRIEL e intenta un posible acuerdo. Antes de la audiencia de juicio, GABRIEL revoca el mandato a ENRIQUE y, auxiliado por un nuevo abogado, llega a un acuerdo con la contraria recibiendo \$7.000.000.- Posteriormente, ENRIQUE demanda a GABRIEL solicitando el pago del 20% de lo obtenido en acuerdo judicial. El tribunal de 1ª instancia y la Corte de Apelaciones rechazan la demanda. La Corte Suprema revoca la sentencia y ordena el pago de lo solicitado. Argumenta que “si se examina la cláusula quinta del convenio antes aludido, se advierte que la participación del profesional en el logro del acuerdo es

un elemento o condición adicional impuesta por los sentenciadores, que no forma parte de lo estipulado. En efecto, la cláusula se limita a señalar que, «en caso que GABRIEL llegue a un acuerdo de carácter judicial o extrajudicial con la demandada, el abogado ENRIQUE recibirá por concepto de honorarios un veinte por ciento de lo que el trabajador acuerde en dicho pacto», de lo que se colige que la única exigencia es que el mandante llegue a ese acuerdo, pero no necesariamente, que lo sea mediado por el profesional que lo asesora. Habiéndose alcanzado un acuerdo entre el demandado y su ex empleador, los jueces debieron aplicar lo convenido por las partes, declarando el derecho del demandante a que le sea pagado, a título de honorarios, el 20% de lo obtenido con el acuerdo”. Agrega que: “el hecho que el demandado le hubiere revocado el patrocinio y poder al demandante antes de realizarse la audiencia preparatoria en que se alcanzó el acuerdo entre las partes, no pone término a la obligación del mandante de remunerar al mandatario en los términos estipulados, en la medida que se ha verificado la circunstancia convenida para que surja la obligación de pagar los honorarios”.

Caso 2. En la cláusula 10^a de un contrato de arrendamiento se dice: “Si cualquiera de las partes del presente contrato decidiere unilateralmente ponerle término a éste por causales distintas a las estipuladas en este contrato, deberá indemnizar al otro contratante con la suma que corresponda pagar como renta de arrendamiento por todo el tiempo que faltare para completar el plazo de seis años fijado para la duración del presente contrato”. Con posterioridad, las partes firman una escritura de terminación de contrato de arrendamiento, en cuya cláusula tercera se dice: “Declara MARÍA, que en conformidad a lo establecido en la cláusula décimo, de la escritura pública de Arrendamiento ya citada precedentemente, que a título de devolución, se obliga a pagar a PEDRO, la suma de quince millones de pesos, pago que hará efectivo cuando celebre un nuevo contrato de arrendamiento, lo que deberá ocurrir a más tardar el día treinta de octubre del año en curso.”

PEDRO demandó a MARÍA el pago de \$15.000.000.-, pues, habiendo vencido este último plazo no le pagó dicha cantidad. MARÍA pidió el rechazo de la demanda dado que la obligación no es actualmente exigible, argumentando que, según la cláusula tercera de la escritura de Terminación, las partes supeditaron la obligación de pago de MARÍA a la celebración de un nuevo contrato de arrendamiento, hecho que hasta la fecha no se ha cumplido. Los tribunales de las instancias rechazaron la demanda al entender que se trataba de una condición suspensiva pendiente. La Corte Suprema revocó el fallo y acogió la demanda (entendió que se trataba de un plazo vencido), indicando que “Si las partes hubiesen querido eximir a la demandada de la obligación resarcitoria, era más razonable que lo expresaran sin preámbulos”.

Caso 3. En un contrato de promesa de compraventa de inmueble se consideran las siguientes cláusulas: i) 5^a “La escritura pública que dé cuenta del contrato de compraventa prometido deberá otorgarse y suscribirse, sin sujeción a modalidad o condición posterior alguna dentro del plazo de cien días hábiles contados desde esta fecha, esto es, a más tardar, el día martes 27 de enero de 2009”; ii) 13^a “La promitente compradora reconoce la intervención de los corredores señores José y Carlos, en todo el proceso de esta compraventa, y se compromete a pagarles una comisión equivalente a un dos por ciento más impuesto al valor agregado del precio total del contrato de compraventa, pago que deberá hacerse efectivo contra firma de escritura definitiva e inscripción del nuevo dominio a nombre de la compradora”. El contrato definitivo no se

suscribió, dado que las partes por medio de una transacción dejaron sin efecto las obligaciones contraídas en el contrato de promesa. La promitente compradora pagó al promitente vendedor, la multa por no celebración del contrato prometido, pero no la comisión a los corredores de propiedad. Los corredores demandaron a la promitente compradora el pago de la comisión alegando que el plazo de la cláusula 13ª en relación a la cláusula 5ª, se encuentra vencido. La demandada, por su parte, sostuvo que nada debía a los corredores pues no se cumplió la condición establecida en la cláusula 13ª, esto es, que se firmara la escritura de compraventa definitiva e inscribiera el dominio a nombre de la compradora (no se trataba de un plazo). La demanda fue acogida en ambas instancias y la Corte Suprema rechazó el recurso de casación en el fondo deducido por la demandada, quedando firme la sentencia de alzada. La Corte Suprema indicó que “a la luz del razonamiento del juez a quo —hecho suyo por el tribunal de alzada— debe entenderse que la interpretación efectuada era la que estimó mejor cuadraba con la naturaleza del contrato de corretaje, por su carácter «asalariado», así como que al interpretar las cláusulas del contrato “unas por otras”, lo que hizo en relación a las cláusulas quinta y décimo tercera del contrato de promesa, de manera de darle el sentido que mejor convenga al contrato en su totalidad, resultaba aplicable el plazo fijado para la conclusión del contrato definitivo”.

Caso 4. Tranex y Anglo celebraron un contrato de carguío y transportes del concentrado extraído en labores mineras. En la cláusula 3.2 se dice: “La compañía Anglo o el contratista Tranex] podrá dar por terminado este contrato en cualquier momento y sin necesidad de invocar causa alguna, bastando para ello que lo comunique por escrito a la otra parte a lo menos con ciento veinte días corridos de anticipación. El ejercicio de esta facultad no dará derecho a ninguna de las partes a indemnización o compensación de ninguna especie”. Según el contrato Tranex se comprometía a transportar material en camión hasta una estación de ferrocarriles, cuya construcción estaba a cargo de otra empresa que nunca lo hizo. Ante dicho inconveniente, Tranex ofreció hacer el traslado completo en camión hasta el Puerto Ventanas, lo que le significó incurrir en sobrecostos de los que Anglo se haría cargo. Mientras se negociaba una modificación al contrato y habiéndose Tranex ya hecho cargo del transporte del concentrado, Anglo la notificó de que ponía término al contrato. Tranex demandó a Anglo alegando de parte de esta mala fe y que debía indemnizarle los cuantiosos costos en que tuvo que incurrir para hacer operativo el transporte del mineral, al asumir el trayecto que inicialmente correspondía hacer por ferrocarril. La Corte Suprema declaró: “Si bien una terminación unilateral anticipada como la de autos debe considerarse como una forma de extinción de los contratos, para evaluar su procedencia es necesario [aclarar] si la parte que se dice perjudicada con esa determinación sufrió perjuicios por el ejercicio inadecuado de tal atribución. Por ende, no parece posible abstraerse sin más de la ulterior responsabilidad que podría conllevar esa unilateral decisión por la sola circunstancia de haberse convenido en el contrato, ya que un término intempestivo, abrupto o abusivo bien podría importar una infracción de la obligación de ejecutar el contrato de buena fe, considerando, como acontece en la especie, que el prestador del servicio pudo incurrir en gastos en su actividad y/o sufrir perjuicios derivados de la resistencia de su contraparte en la continuación del vínculo que los ligaba, todo lo cual hace razonable reconocer el derecho a la indemnización de todo el daño que se sufra como consecuencia del término de los servicios”

Ahora Chat GPT tengo que informarte que para tomar decisiones la Corte Suprema tiene que considerar especialmente los siguientes artículos del Código Civil:

Art. 1545. Todo contrato legalmente celebrado es una ley para los contratantes, y no puede ser invalidado sino por su consentimiento mutuo o por causas legales.

Art. 1546. Los contratos deben ejecutarse de buena fe, y por consiguiente obligan no sólo a lo que en ellos se expresa, sino a todas las cosas que emanan precisamente de la naturaleza de la obligación, o que por la ley o la costumbre pertenecen a ella.

Art. 1560. Conocida claramente la intención de los contratantes, debe estarse a ella más que a lo literal de las palabras.

Art. 1561. Por generales que sean los términos de un contrato, sólo se aplicarán a la materia sobre que se ha contratado.

Art. 1562. El sentido en que una cláusula puede producir algún efecto, deberá preferirse a aquel en que no sea capaz de producir efecto alguno.

Art. 1563. En aquellos casos en que no apareciere voluntad contraria deberá estarse a la interpretación que mejor cuadre con la naturaleza del contrato. Las cláusulas de uso común se presumen aunque no se expresen.

Art. 1564. Las cláusulas de un contrato se interpretarán unas por otras, dándose a cada una el sentido que mejor convenga al contrato en su totalidad. Podrán también interpretarse por las de otro contrato entre las mismas partes y sobre la misma materia. O por la aplicación práctica que hayan hecho de ellas ambas partes, o una de las partes con aprobación de la otra.

Art. 1565. Cuando en un contrato se ha expresado un caso para explicar la obligación, no se entenderá por sólo eso haberse querido restringir la convención a ese caso, excluyendo los otros a que naturalmente se extienda.

Art. 1566. No pudiendo aplicarse ninguna de las reglas precedentes de interpretación, se interpretarán las cláusulas ambiguas a favor del deudor. Pero las cláusulas ambiguas que hayan sido extendidas o dictadas por una de las partes, sea acreedora o deudora, se interpretarán contra ella, siempre que la ambigüedad provenga de la falta de una explicación que haya debido darse por ella.

Ahora viene el caso que me interesa que resuelvas si piensas de la forma en que lo ha hecho la Corte Suprema en los casos que te indiqué:

Eldu demandó (i) la indemnización de los perjuicios patrimoniales y extrapatrimoniales que Eletrans le habría causado como consecuencia de diversos incumplimientos de un Contrato y (ii) la restitución o pago de ciertos bienes que alega habrían sido de su propiedad pero que se encontrarían en posesión de Eletrans.

Eldu fundó su demanda indemnizatoria en los siguientes incumplimientos imputados a Eletrans.

A) En la Cláusula 5.02.01. del documento “Bases de Pago Contrato STT1-CDA-001/2016” los trabajos debían pagarse de acuerdo con el desarrollo o avance de las obras, de manera proporcional al precio pactado en la suma alzada.

B) Al 19 de enero de 2018, Eldu había ejecutado el 70,19% de los trabajos encomendados, cumpliendo el hito constructivo N°9 establecido en el Contrato (“Hito N°9”), pero habría

recibido el pago únicamente por el equivalente al 20% del avance de las obras, quedando pendiente el pago del 50,19% del precio del Contrato.

C) Eletrans dio término anticipado al Contrato, de manera arbitraria, infundada y abusiva.

Eldu reclamó el pago de la siguiente cantidad:

Los costos en que habría incurrido por la ejecución del 70,19% de las obras, descontándose lo ya pagado en virtud de los avances cubiertos por los EDP Nos1 y 2. El monto reclamado fue de \$616.503.588.

Una indemnización de \$300.000.000. por los perjuicios que le provocó la terminación anticipada del contrato, de manera arbitraria, infundada y abusiva.

Eletrans contestó la demanda interpuesta en su contra indicando que:

El sistema de pago convenido establecía la obligación de realizar los pagos en función del cumplimiento de hitos pactados bajo el Contrato y no respecto del estado de avance de las obras. Fundamentó lo anterior en lo dispuesto en la Cláusula 5.03. en la que se dice

44) La Demandada sostuvo haber pagado en tiempo y forma los estados de pago correspondientes a los hitos de pago efectivamente cumplidos, esto es, los Estados de Pago Nos1 y 2.

45) Eletrans explicó que no pagó el estado de pago N°3 (“EDP N°3”) presentado por Eldu el 30 de enero de 2018, ya que esta no había acompañado toda la documentación exigida por el Contrato para la aprobación del estado de pago.

B. Contestación y dúplica de Eletrans

Respecto de los incumplimientos imputados por Eldu

Eletrans pagó efectivamente los trabajos ejecutados por Eldu. De acuerdo a lo establecido en el contrato el pago en caso de terminación anticipada de contrato, se consideraría el valor resultante del cumplimiento de ciertos hitos, lo cual coincide con el estado de pago N°3. En ninguna parte se indica que se debe pagar según estado de avance.

Prueba presentada

Se dio por probado que el estado de pago n°3 consideraba hasta el hito 9, lo que equivale a \$350.000.000.- y que no ha sido pagado a Eldu.

Se dio por probado que el estado de avance de la obra es de 70,19%, siendo el valor de la diferencia con lo ya pagado de un monto de \$616.503.588.-

Se dio por probado que según cláusula del contrato Eletrans podía poner término anticipado al contrato sin expresión de causa y que solo estaría obligada a realizar los pagos pendientes, según estados de pago calculado conforme a cumplimiento de hitos.

Se dio por probado que si no se hubiera dado término anticipado al contrato, Eldu habría tenido una ganancia líquida de \$300.000.000.-

¿Cuál sería el monto que la Corte Suprema debiera ordenar pagar a Eldu y qué tendría que decir la sentencia?

Al recibir las evaluaciones de resultados por parte de los abogados, resultó ser que las mejores respuestas, fueron las del modelo GPT-4 y las que contaban de mayor contexto previo, en base a esto, se escogieron las mejores dos respuestas de cada caso. A continuación, se muestran las respuestas escogidas.

- Primer caso [Respuesta 1]

La situación planteada es compleja y tiene elementos tanto contractuales como de interpretación de las cláusulas dentro del contrato. Sin embargo, considerando la legislación pertinente en Chile y los antecedentes expuestos, aquí está la resolución que podría tomar la Corte Suprema:

I. Antecedentes: Los hechos son claramente expuestos en la presentación del caso. Las partes celebraron un contrato que contenía una cláusula que permitía a Cimientos Industriales José Manuel rescindir el contrato con preaviso de 30 días, sin derecho a indemnización para Los Palotes Limitada. Posteriormente, Cimientos Industriales José Manuel optó por ejercer esta facultad.

II. Consideraciones:

1. Validez de la Cláusula: El contrato fue aceptado por ambas partes, y la cláusula de rescisión anticipada fue incluida en él. La empresa Los Palotes Limitada tenía conocimiento de dicha cláusula y aceptó los términos. La cláusula no parece ser contraria a la ley o la moral, y no hay evidencia de que haya sido impuesta de manera abusiva.

2. Buena Fe Contractual: Aunque Los Palotes Limitada argumenta que la cláusula de rescisión anticipada sin motivo atenta contra la buena fe contractual, el hecho de que ambas partes acordaran dicha cláusula y que no haya evidencia de conducta desleal por parte de Cimientos Industriales José Manuel parecería contrarrestar este argumento.

3. Intención de las Partes y Naturaleza del Contrato: A pesar de que Los Palotes Limitada pueda haber incurrido en costos e inversiones significativas, la posibilidad de rescisión anticipada estaba claramente establecida en el contrato. La intención de las partes debe ser determinada por lo que se ha pactado expresamente.

4. Ejercicio de la Facultad de Rescindir: Cimientos Industriales José Manuel parece haber ejercido la facultad de rescindir el contrato de conformidad con lo establecido en el contrato, incluyendo el preaviso requerido.

5. Quantum Demandado y Perjuicios: La demanda de indemnización por parte de Los Palotes Limitada no parece tener fundamento en la medida en que el contrato excluía expresamente el derecho a indemnización en caso de rescisión anticipada.

III. Fallo: En consecuencia, y en consideración a los antecedentes expuestos y la legislación aplicable, se rechaza la demanda presentada por Los Palotes Limitada en cuanto a la indemnización de perjuicios y cumplimiento del contrato. Se determina que

la rescisión del contrato por parte de Cimientos Industriales José Manuel fue ejercida conforme a los términos pactados en el contrato, y que la cláusula de rescisión anticipada es válida y aplicable.

La decisión sobre las costas del juicio podría depender de una evaluación más detallada de la conducta procesal de las partes.

Este fallo se basa en la información proporcionada y en la legislación chilena pertinente. Cabe señalar que la interpretación final de un caso de esta naturaleza podría depender de factores adicionales no contemplados aquí, y correspondería a un tribunal competente analizar todas las pruebas y argumentos presentados por las partes.

- Primer caso [Respuesta 2]

Asumiendo el papel de la Corte Suprema de Chile y tomando en consideración los principios y reglas aplicables al caso, sin mencionar explícitamente los artículos indicados anteriormente, procedería a resolver la controversia de la siguiente manera:

1. Validez de la Cláusula de Rescisión Anticipada: En el contrato existente entre las partes, se estableció una cláusula que permitía a Cimientos Industriales José Manuel rescindir el contrato anticipadamente, sin la necesidad de alegar una razón, y sin que Los Palotes Limitada pudiera reclamar indemnización o pagos adicionales. Esta cláusula estaba claramente expresada en el contrato, y Los Palotes Limitada aceptó sus términos al firmarlo.

2. Interpretación del Contrato: La interpretación de los contratos debe hacerse de buena fe, dando preferencia al sentido que mejor convenga a la naturaleza del contrato y a la intención de las partes. Si bien Los Palotes Limitada sostiene que la cláusula es anacrónica y vulnera la buena fe, la cláusula fue clara, y no hay evidencia de ambigüedad o contradicción en su redacción. Ambas partes son empresas comerciales que, se presume, actuaron con conocimiento y comprensión de los términos del contrato.

3. Ejercicio de la Facultad de Rescisión: Cimientos Industriales José Manuel ejerció su derecho de rescindir el contrato de acuerdo con los términos pactados. Aunque la rescisión anticipada pueda haber causado perjuicios a Los Palotes Limitada, la cláusula de rescisión anticipada fue clara y explícita en cuanto a que no habría derecho a indemnización.

4. Bases de la Licitación: Además, la cláusula estaba contenida en las bases de la licitación, y Los Palotes Limitada realizó su oferta sin pronunciarse sobre ella. La empresa también había participado en otras licitaciones con cláusulas similares, lo que refuerza la idea de que aceptaron los términos de manera informada.

5. Buena Fe y Equilibrio Contractual: No se aprecia un desequilibrio contractual ni una violación de la buena fe que haga inaplicable la cláusula. La posibilidad de rescisión anticipada fue un término negociado y acordado por ambas partes.

6. Conclusión: En vista de lo anterior, la demanda de Los Palotes Limitada sería desestimada, y se confirmaría la validez de la rescisión anticipada del contrato por parte de Cimientos Industriales José Manuel. La cláusula fue clara, las partes estaban en condiciones de comprenderla, y se ejerció de acuerdo con los términos del contrato.

Este fallo busca respetar los principios de autonomía contractual, cumplimiento de los contratos, buena fe, y la necesidad de claridad y certeza en los acuerdos comerciales. Los principios legales relevantes aplicados aquí reflejan la necesidad de que los contratos sean vinculantes y de que las partes estén obligadas a cumplir con los términos acordados, a menos que haya razones legales para invalidar o modificar esos términos.

- Segundo caso [Respuesta 1]

Basándome en el resumen que has proporcionado y en los principios legales generalmente aplicados, el razonamiento de la Corte Suprema podría seguir las siguientes líneas:

El primer punto a resolver es el incumplimiento del pago por parte de Eletrans. De acuerdo con las pruebas presentadas, Eldu cumplió con el hito N°9, pero solo recibió pago por el 20% del avance de las obras, quedando pendiente el pago del 50,19% del precio del contrato. Según el estado de pago n°3, este monto asciende a \$350.000.000.

Por otro lado, Eletrans alega que su sistema de pago se basa en el cumplimiento de hitos y no en el estado de avance de las obras. Sin embargo, el estado de pago n°3 (que representa el hito N°9) no fue pagado, a pesar de haber sido cumplido por Eldu. En cuanto a la terminación anticipada del contrato, aunque Eletrans tenía el derecho de terminarlo sin expresión de causa, debía pagar los costos pendientes. La ganancia líquida que Eldu habría obtenido de no ser por la terminación anticipada fue de \$300.000.000.

Por lo tanto, la Corte Suprema podría dictaminar lo siguiente:

1. *Eletrans debe pagar a Eldu los \$350.000.000 correspondientes al estado de pago n°3 (hito N°9).*

2. *No procede la indemnización de \$300.000.000 demandada por Eldu por perjuicios debido a la terminación anticipada del contrato. La terminación anticipada estaba permitida por el contrato y no hay indicios de que se haya hecho de manera abusiva. Sin embargo, Eletrans debía cumplir con sus obligaciones contractuales pendientes al momento de la terminación, por lo que debe pagar el estado de pago n°3.*

Esto asume que los términos del contrato no proporcionan otras vías para calcular los daños o indemnizaciones debidos a Eldu.

Esto es solo una interpretación y un posible resultado. Los detalles exactos y el resultado final dependerían de los términos específicos del contrato y de la jurisprudencia aplicable.

- Segundo caso [Respuesta 2]

Según los antecedentes proporcionados y tomando en cuenta los principios aplicados por la Corte Suprema en casos similares, aquí está mi análisis:

1. *Respecto a los trabajos realizados: La discrepancia parece estar en la interpretación de las cláusulas del contrato relacionadas con los pagos. Eldu interpreta que debe ser pagado de acuerdo con el progreso general del proyecto (70,19% en este caso), mientras que Eletrans interpreta que los pagos deben realizarse según el cumplimiento de hitos específicos. Dado que las pruebas confirmaron que Eldu completó hasta el hito N°9 y que esto equivale a \$350.000.000 que no han sido pagados, la Corte Suprema*

probablemente ordenaría a Eletrans pagar a Eldu esta cantidad. Esto se basa en el cumplimiento de los hitos y no en el avance general del trabajo.

2. Respecto a la terminación anticipada del contrato: Según la cláusula contractual, Eletrans tiene derecho a terminar el contrato anticipadamente sin necesidad de justificar la causa y sólo se le requeriría pagar los hitos cumplidos pendientes. En este caso, eso sería el monto asociado con el hito N°9 (\$350.000.000).

3. Respecto a la indemnización solicitada: La Corte Suprema podría negar la indemnización de \$300.000.000 solicitada por Eldu por la terminación anticipada del contrato. Si bien Eldu afirma que habría obtenido una ganancia líquida de \$300.000.000 si el contrato no hubiera sido terminado anticipadamente, la cláusula contractual permite a Eletrans terminar el contrato anticipadamente sin expresar la causa. Como tal, Eldu tendría que demostrar que la terminación fue en realidad "arbitraria, infundada y abusiva" para obtener algún tipo de indemnización, lo que parece ser un argumento difícil de sostener dado el derecho contractual de Eletrans.

En resumen, basándose en la interpretación de las cláusulas del contrato y la evidencia presentada, la Corte Suprema probablemente ordenaría a Eletrans pagar a Eldu los \$350.000.000 correspondientes al cumplimiento del hito N°9 y rechazaría la indemnización solicitada por Eldu por la terminación anticipada del contrato. La sentencia, por lo tanto, enfatizaría la interpretación de las cláusulas contractuales y la necesidad de adherirse a ellas, así como el derecho de Eletrans a terminar el contrato anticipadamente según lo establecido en el contrato.

4.2 Preparación de los datos

De acuerdo con la guía que ofrece OpenAI [] para el fine-tuning de sus modelos, es necesario ingresar los ejemplos en un formato JSONL siguiendo la siguiente estructura de ejemplo:

```
1 {"messages": [{"role": "system", "content": "Marv is a factual chatbot that is also sarcastic."}, {"role": "user", "content": "What's the capital of France?"}, {"role": "assistant", "content": "Paris, as if everyone doesn't know that already."}]}
```

```
2 {"messages": [{"role": "system", "content": "Marv is a factual chatbot that is also sarcastic."}, {"role": "user", "content": "Who wrote 'Romeo and Juliet'?"}, {"role": "assistant", "content": "Oh, just some guy named William Shakespeare. Ever heard of him?"]}
```

```
3 {"messages": [{"role": "system", "content": "Marv is a factual chatbot that is also sarcastic."}, {"role": "user", "content": "How far is the Moon from Earth?"}, {"role": "assistant", "content": "Around 384,400 kilometers. Give or take a few, like that really matters."}]}
```

En base a esto es que se completó manualmente la estructura, contemplando como contenido de usuario los casos presentados sin contexto previo y como contenido de asistente las mejores respuestas, con el objetivo de conseguir las mejores respuestas con el menor contexto posible, agregando en esta ocasión, el siguiente contenido de sistema: *“Eres un consultor legal chileno especializado en interpretación judicial de contratos.”* También fue necesario especificar manualmente los saltos de línea con el siguiente texto `“\n”`. Sumado a esto se triplicó cada ejemplo para lograr un total de 12 ejemplos, ya que OpenAI exige un mínimo de 10.

Adicionalmente para corregir errores de formato, es que se utilizó un extenso código provisto por la misma guía de OpenAI para ello, gracias a esto, adicionalmente se identificó que se cuenta con 23 550 tokens en total, y que se estima que se entrenará con 8 épocas, cobrándose en total por 188 400 tokens, lo cual resultó ser equivalente a \$1.51 USD

4.3 Modelado

Para acceder a la API proporcionada por OpenAI para realizar el fine-tuning es necesario contar con una llave personal, que es adquirida creándola desde los ajustes de la cuenta. De esta manera la API puede reconocer a que cuenta asociar los datos, los modelos y los cobros de uso respectivos.

Luego es necesario ejecutar el siguiente código para crear el archivo con los ejemplos construidos:

```
1 import openai
2 import os
3 openai.api_key = os.getenv("OPENAI_API_KEY")
4
5 openai.File.create(
6     file=open("mydata.jsonl", "rb"),
7     purpose='fine-tune'
8 )
```

Luego para crear y agregar el trabajo de fine-tuning a la lista de espera de OpenAI, es necesario ejecutar el siguiente comando:

```
1 openai.FineTuningJob.create(training_file=file_id, model="gpt-3.5-turbo")
```

Cuando se termina el proceso, que no suele tardar más de 10 minutos, se recibe un correo a la cuenta donde se proporciona el id del modelo en cuestión, el cual será necesario si se desea utilizar aún en formato de código para una aplicación personal, en caso contrario, se puede utilizar la sección web Playground que facilitan para ello.

4.4 Evaluación

Para evaluar la calidad de las respuestas generadas por el nuevo modelo, se creó un documento donde se colocaron 3 respuestas del modelo entrenado y 3 respuestas del modelo sin entrenar, para los dos casos.

El profesor Rodrigo Coloma logró reconocer que en definitiva el entrenamiento fue exitoso, y el modelo entrenado es capaz de resumir bien la manera de lidiar con el problema y tomar una decisión, a diferencia del modelo sin entrenar que es más reacio a tomar una decisión y se limita en mayor parte a recapitular la información presentada.

A modo de ilustración, se presentan dos respuestas generadas por el modelo entrenado en comparación al no entrenado:

- Primer Caso [Respuesta Modelo Entrenado]

Asumiendo que no hay disposiciones legales que modifiquen el contrato o que sean aplicables al caso, y considerando los principios generales del derecho contractual en Chile, así como la jurisprudencia pertinente, la resolución de la controversia podría seguir los siguientes lineamientos:

1. *Validez de la Cláusula:* La cláusula que permite a Cimientos Industriales José Manuel rescindir el contrato con preaviso de 30 días, sin derecho a indemnización para Los Palotes Limitada, está contenida en el contrato y fue aceptada por ambas partes. La empresa Los Palotes Limitada, al aceptar dicha cláusula, asumió el riesgo de una rescisión anticipada y renunció a su derecho a reclamar indemnización.

2. *Buena Fe Contractual:* Aunque Los Palotes Limitada argumenta que la cláusula de rescisión anticipada sin motivo atenta contra la buena fe contractual, la existencia de la cláusula en el contrato y la aceptación de la misma por ambas partes indican que se trató de un acuerdo voluntario. La buena fe debe ser interpretada dentro del marco de lo pactado por las partes.

3. *Intención de las Partes y Naturaleza del Contrato:* A pesar de que Los Palotes Limitada pueda haber incurrido en costos e inversiones significativas, la posibilidad de rescisión anticipada estaba claramente establecida en el contrato. La intención de las partes debe ser determinada por lo que expresamente acordaron, y la naturaleza del contrato no parece ser incompatible con la cláusula de rescisión.

4. *Ejercicio de la Facultad de Rescindir:* Cimientos Industriales José Manuel ejerció la facultad de rescindir el contrato de acuerdo con los términos establecidos en el contrato, incluyendo el preaviso requerido. El hecho de que la rescisión se haya realizado sin motivo no parece contravenir el contrato, dado que la posibilidad de rescisión sin motivo estaba prevista.

5. *Interpretación de las Cláusulas:* La interpretación de las cláusulas contractuales debe hacerse de manera objetiva, atendiendo a su tenor literal y al contexto en que fueron acordadas. No parece haber ambigüedad en la cláusula de rescisión anticipada.

6. *Quantum Demandado y Perjuicios:* La demanda de indemnización por parte de Los Palotes Limitada parece estar en contradicción con la cláusula que excluye el derecho a reclamar indemnización. Los perjuicios alegados, como el daño emergente, lucro cesante y daño moral, podrían no ser procedentes en virtud de la renuncia expresa a indemnización.

7. *Principio Pacta Sunt Servanda:* El principio de cumplimiento de los contratos (pacta sunt servanda) es fundamental en el derecho contractual, y las partes deben cumplir con lo pactado, a menos que haya razones legales para modificar o anular el contrato.

En conclusión, a falta de disposiciones legales que modifiquen el contrato o que sean aplicables al caso, y considerando los principios generales del derecho contractual, parece que la validez de la cláusula de rescisión anticipada, la aceptación de dicha cláusula por parte de ambas partes, y el ejercicio de la facultad de rescindir el contrato de acuerdo con los términos establecidos en el contrato, son elementos que favorecerían a Cimientos Industriales José Manuel en la resolución de la controversia. Sin embargo, la decisión final correspondería a la Corte Suprema de Chile, tomando en cuenta la legislación pertinente y la jurisprudencia aplicable.

- Primer Caso [Respuesta Modelo GPT-3.5-Turbo]

En base a los antecedentes presentados, la controversia entre las partes se centra en la interpretación y validez de la cláusula que permite a Cimientos Industriales José Manuel

rescindir el contrato de forma anticipada, sin necesidad de acciones judiciales ni indemnización, en caso de incumplimiento por parte de Los Palotes Limitada.

La empresa demandante argumenta que esta cláusula es anacrónica, carece de aplicabilidad y debe considerarse no escrita. Sostiene que su ejercicio por parte de Cimientos Industriales José Manuel vulnera la buena fe contractual, altera el equilibrio y la reciprocidad del contrato, y genera perjuicios económicos y morales.

Por su parte, la demandada alega que la cláusula es válida, ya que estaba incluida en las bases de la licitación y Los Palotes Limitada no se pronunció al respecto al presentar su oferta. Además, señala que la empresa demandante ya había celebrado otros contratos con cláusulas similares en licitaciones anteriores.

En este contexto, corresponde a la Corte Suprema de Chile resolver la controversia y determinar la validez y aplicabilidad de la cláusula en cuestión. Para ello, deberá analizar la redacción y los términos del contrato, así como la intención de las partes al momento de celebrarlo. También deberá considerar la naturaleza del servicio prestado, los antecedentes de licitaciones anteriores y cualquier otra circunstancia relevante para la interpretación del contrato.

En caso de que la Corte Suprema determine que la cláusula es válida y aplicable, deberá evaluar si Cimientos Industriales José Manuel ejerció correctamente su facultad de rescindir el contrato y si existen motivos suficientes para ello. En caso de que se determine que la rescisión fue injustificada, la empresa demandante podría tener derecho a reclamar indemnización por los perjuicios sufridos.

En cambio, si la Corte Suprema considera que la cláusula es inválida o inaplicable, deberá evaluar los perjuicios sufridos por Los Palotes Limitada como consecuencia de la rescisión anticipada del contrato y determinar si corresponde otorgar una indemnización.

En resumen, la Corte Suprema de Chile deberá resolver la controversia entre las partes, determinando la validez y aplicabilidad de la cláusula de rescisión anticipada del contrato y evaluando los perjuicios sufridos por la empresa demandante en caso de que la rescisión sea considerada injustificada.

- Segundo Caso [Respuesta Modelo Entrenado]

Según los antecedentes proporcionados y considerando los principios legales aplicables, el razonamiento de la Corte Suprema podría seguir las siguientes líneas:

1. *Contrato y Cumplimiento: La discrepancia entre las partes parece residir en la interpretación de las cláusulas del contrato relacionadas con los pagos y el avance de las obras. Eldu interpreta que debe ser pagado proporcionalmente al avance total del trabajo (70,19%), mientras que Eletrans interpreta que debe pagar por hito cumplido, en este caso hasta el hito N°9.*

2. *Término Anticipado del Contrato: Las cláusulas contractuales permiten a Eletrans terminar el contrato anticipadamente sin necesidad de justificación y pagar solo los hitos cumplidos pendientes. Esto significa que, según la interpretación de Eletrans, deben pagar por el hito N°9 (equivalente a \$350.000.000).*

3. *Demanda de Indemnización:* El monto reclamado por Eldu para la indemnización de los perjuicios (\$616.503.588) parece estar basado en su interpretación de los pagos por avance y en los costos incurridos hasta el 70,19% de avance. Sin embargo, si el contrato permite el término anticipado sin causa y el beneficio esperado para Eldu habría sido de \$300.000.000, entonces la indemnización solicitada parece excesiva.

4. *Fallo Propuesto:* En vista de las cláusulas contractuales y las pruebas presentadas, la Corte Suprema podría determinar lo siguiente:

-Eletrans debe pagar a Eldu el monto correspondiente al cumplimiento del hito N°9 (\$350.000.000).

-La indemnización solicitada por Eldu de \$300.000.000 por terminación anticipada del contrato de manera arbitraria podría ser rechazada, ya que el contrato permite el término anticipado sin expresión de causa.

-La sentencia podría hacer hincapié en la importancia de la interpretación de las cláusulas contractuales y la necesidad de cumplirlas, así como en la necesidad de claridad y comunicación efectiva entre las partes para evitar conflictos en la interpretación.

Este análisis se basa en los antecedentes y principios legales proporcionados, y la resolución final podría depender de otros factores no mencionados o de la jurisprudencia aplicable.

- Segundo Caso [Respuesta Modelo GPT-3.5-Turbo]

Como consultor legal especializado en interpretación judicial de contratos, puedo analizar el caso y proporcionar una opinión basada en la información proporcionada.

Según los argumentos presentados por Eldu y Eletrans en el caso, la disputa se centra en la interpretación del contrato y los pagos correspondientes al cumplimiento de los hitos pactados. Eldu alega que se le debe el pago por el 50,19% restante de los trabajos realizados y reclama una indemnización por la terminación anticipada del contrato, mientras que Eletrans sostiene que solo está obligada a pagar según los hitos pactados y que Eldu no ha cumplido con los requisitos para recibir el pago correspondiente al estado de pago N°3.

Teniendo en cuenta los elementos presentados en la demanda, contestación y réplica, así como las pruebas presentadas, la Corte Suprema deberá examinar cuidadosamente la Cláusula 5.02.01 y 5.03 del contrato para determinar si los pagos deben realizarse de acuerdo con el estado de avance de las obras o según los hitos pactados.

Si la Corte Suprema determina que los pagos deben realizarse según el estado de avance de las obras, entonces Eldu tendría derecho a recibir el pago correspondiente al 50,19% restante de los trabajos realizados, equivalentes a \$616.503.588. Además, si la terminación anticipada del contrato por parte de Eletrans se considera arbitraria, infundada y abusiva, la Corte podría ordenar una indemnización adicional de \$300.000.000.

Sin embargo, si la Corte Suprema determina que los pagos deben realizarse según los hitos pactados, entonces Eletrans podría argumentar que ha cumplido con su obligación

de pago, ya que solo ha pagado los estados de pago Nos 1 y 2 correspondientes a los hitos cumplidos.

En última instancia, la sentencia de la Corte Suprema dependerá de cómo interprete los términos del contrato y la evidencia presentada por ambas partes.

Finalmente, en palabras textuales, se presenta la evaluación final por parte del profesor Rodrigo Coloma: *“La principal diferencia entre las respuestas es que el modelo no entrenado se limita a reconstruir lo que se le dice (no decide). El modelo entrenado, en cambio, sí toma una decisión (por supuesto es muy breve, pero suficientemente estructurada). En el primer caso muy similar a la que efectivamente dictó la Corte Suprema. En el segundo caso la respuesta es similar a la que adoptó el tribunal arbitral de 2a instancia (el fallo de la Corte Suprema se aleja, en parte, a lo que ella venía diciendo en otras sentencias (lo que hace Chat GPT, entonces, es bastante consistente con lo que se le enseñó: no asume riesgos).”*

Capítulo 5

Análisis de Resultados

En este capítulo se estudia el por qué del fracaso de las alternativas desechadas, y lo que les falta para llegar a conseguir mejores resultados, luego se analizan las implicancias de los resultados conseguidos con el modelo final, y finalmente se ofrecen sugerencias para mejorar los resultados futuros.

5.1 Alternativas desechadas

Una de las alternativas era analizar o interpretar el contrato en si mismo, buscando patrones de error que se pudieran encontrar en un vasto numero de ejemplares, el problema principal empieza en su adquisición, ya que, al ser de carácter privado, se complica la idea de conseguir una cantidad considerable de contratos que hayan sido sometidos a juicio y que adicionalmente sean lo suficientemente parecidos para extraer patrones de estos. Aún si fuese posible conseguir esta cantidad, surge la dificultad de interpretar semejantes cantidades de datos utilizando recursos computacionales reducidos, ya que los modelos que tienen esa capacidad requieren de costosos servidores dedicados.

Otra de las alternativas era utilizar los documentos de demanda, contestación y sentencia, dicha alternativa comparte la dificultad de interpretar semejante cantidad de datos utilizando recursos computacionales no dedicados, y adicionalmente, se le agrega el problema del formato en que estos se encuentran muchas veces, donde muchos de estos al ser documentos escaneados requieren de programas de reconocimiento óptico de caracteres (OCR) para la extracción de su contenido textual, donde si bien se han conseguido avances importantes en el área, se pudo observar que aún requieren de una significativa intervención humana para corregir detalles significativos.

Otra de las alternativas era utilizar modelos Open Source para conseguir un modelo gratuito, sin embargo, particularmente en el área legal, estos modelos poseen una capacidad de interpretación muy por debajo de los estándares conseguidos por ChatGPT, ahora bien, esto hace sentido cuando se toma en cuenta que no es un área particularmente popular y de libre acceso, por lo que no suelen tomar este tipo de contenido en los entrenamientos de estos modelos, por lo que sería necesario que alguien no sólo decidiera invertir en costosos recursos computacionales para generar este tipo de entrenamiento masivo, si no que además debiera tener acceso a un vasto contenido legal para poder aspirar a conseguir resultados ejemplares.

La última alternativa fallida fue la de realizar un ajuste fino al modelo davinci de OpenAI, que consiste en un modelo que no ha sido entrenado para ser un chatbot por lo que no contempla varias de las funcionalidades para comprender e interpretar instrucciones que tiene ChatGPT, por lo mismo, es que hace sentido que los resultados de su entrenamiento, contemplando un uso tan reducido de ejemplos, tan sólo haya conseguido una generación de texto poco coherente, además este modelo en particular tan sólo puede lidiar con 2048 tokens entre entrada y salida, lo cual es demasiado reducido para las necesidades de este proyecto, en comparación con los 4096 tokens que puede lidiar el nuevo modelo gpt-3.5-turbo, que adicionalmente está pre entrenado con las características para ser un chatbot.

5.2 Implicancias de los resultados

Obtener resultados tan destacados con tan pocos ejemplos es una hazaña notable, especialmente considerando que, generalmente, se requieren al menos 100 ejemplos para empezar a percibir diferencias significativas. Sin embargo, en este caso específico, el modelo funcionó excepcionalmente bien, resaltando sus asombrosas capacidades para este tipo de proyectos. Además, es digno de mencionar el corto período de tiempo necesario para realizar el ajuste fino (fine-tuning) de un modelo tan complejo y con resultados tan prometedores. Junto con su facilidad de acceso y sus bajos costos de entrenamiento y uso, indudablemente es una herramienta que promete abrir camino para futuros proyectos más ambiciosos que aspiren a superar los límites conocidos y elevar la calidad de las respuestas al siguiente nivel.

Es importante considerar ciertos aspectos sobre la calidad de las respuestas. En primer lugar, las respuestas óptimas solo se extrajeron del modelo GPT-4 con un cierto grado de ingeniería de consulta para maximizar su rendimiento, pero no se realizó ninguna modificación a la respuesta en sí. Esto significa que, hasta ahora, solo se ha facilitado el acceso a las mejores respuestas que el modelo GPT-4 puede proporcionar, en lugar de redactar la mejor respuesta posible desde cero o con la ayuda de estos modelos. Esto está en consonancia con el objetivo inicial del proyecto, que era determinar la posibilidad y el alcance actual de la tecnología. Sin embargo, para intentar mejorar o evaluar de manera más efectiva los resultados, se enfrenta a una tarea particularmente complicada, dado que el vocabulario y estilo son tan específicos que resulta difícil para alguien no experto en el tema definir o reconocer lo que es bueno o no. Incluso para un experto, es especialmente complejo construir una consulta lo suficientemente rigurosa y concisa para incluirla en el modelo y, además, estudiar la rigurosidad de la respuesta.

Es importante tener en cuenta que, aunque una respuesta haya sido validada por un tribunal, no necesariamente significa que sea la correcta. La Corte de Apelaciones o la Corte Suprema podrían revocar dicha sentencia, lo que plantea la posibilidad de que solo lo que dictamine la Corte Suprema sea considerado como la respuesta correcta. Sin embargo, incluso esto es objeto de debate, ya que las respuestas de la Corte Suprema no siempre son consistentes. Por lo tanto, resulta especialmente interesante analizar qué tipo de respuesta podría generar una inteligencia artificial, la cual se presume imparcial, aunque podría tener sus propios sesgos derivados del entrenamiento o de las políticas de la empresa que la desarrolló.

5.3 Recomendaciones para el futuro

En caso de seguir con esta misma línea de trabajo, es necesario tener en consideración que al ritmo que están avanzando los modelos LLM, es probable que dentro de poco se liberen herramientas que superen todas las limitaciones y alcances que se tienen hoy, por dar un ejemplo, OpenAI ya tiene programado la liberación de los modelos para el ajuste fino gpt-3.5-turbo-16k y gpt-4 para inicios del 2024. Teniendo en consideración estas noticias, es que las recomendaciones son a seguir avanzando en la metodología para conseguir los mejores pares de pregunta respuesta posibles, la primera diferencia radicará en la cantidad de contexto que pueden interpretar a la vez estos modelos, lo que da espacio para aspirar a respuestas más detalladas y rigurosas, por lo que a su vez, es necesario contar con un estudio aún más completo y detallado de las preguntas y contexto que se deben incluir.

Es necesario entonces, en próximas ocasiones, contar con un equipo más amplio de abogados para la construcción de preguntas estructuradas y capacidad de análisis exhaustivo de las respuestas, de manera de poder iterar de forma más rigurosa hacia respuestas más profesionales, e idealmente mayor cantidad de ejemplos para conseguir un entrenamiento con menor riesgo de caer en el problema del sobre ajuste, es decir, que sólo sepa resolver los casos con los que se entrenó.

Es importante considerar, que al tratarse de una máquina entrenada para dar la respuesta más aceptable a nivel macro, es particularmente susceptible a dar respuestas genéricas si es que no se le da una instrucción lo suficientemente específica, lo cual entonces también habría que considerar para no darle contexto o instrucciones que puedan incitarle a tomar una postura o tendencia en particular, que pudiera ser considerado sesgo.

Otro punto que podría ser desarrollado en investigaciones futuras sería la identificación de distintos trasfondos desde los que se producen las decisiones de la Corte Suprema. Así, si se observa los casos seleccionados para alimentar a Chat GPT se puede concluir que los valores que se están protegiendo son diferentes. En tal sentido, es preciso un trabajo colaborativo entre expertos de distintas disciplinas a los efectos de identificar los factores que influyen en la Corte Suprema para operar desde un trasfondo u otro.

Capítulo 6

Conclusiones

Gracias al trabajo realizado, se pudo apreciar que se cumplieron las hipótesis de investigación, es decir, es factible con la tecnología actual, rescatar el sentido semántico de oraciones judiciales chilenas de gran complejidad, identificar estructuras tipo de demandas o sentencias chilenas, y de entrenar un modelo que genere respuestas similares a la Corte Suprema, a conflictos de interpretación judicial de contratos, cumpliendo así los objetivos del proyecto.

En el presente trabajo se logró explorar de manera exhaustiva el alcance tanto teórico como práctico de la tecnología vigente en el área de procesamiento de lenguaje natural aplicado al área judicial, se exploraron diversas alternativas de acercamiento y se detallaron las razones por las cuales no son convenientes, también se presentó una metodología diseñada para evaluar la calidad de las respuestas generadas por diferentes modelos, y se reporta empíricamente la pequeña brecha existente entre el mejor modelo LLM de la actualidad y el sistema judicial chileno.

Entre los hallazgos más significativos, se encuentra el avanzado conocimiento actual de los modelos de OpenAI sobre la estructura y redacción tipo del área judicial chileno, dando una idea de la inmensa base de datos con la que fueron entrenados para llegar a tener semejante dominio de un área tan particular. También es relevante destacar las capacidades del reciente lanzamiento de su modelo tipo chat para ajuste fino gpt-3.5-turbo, que es sin duda un paso tremendo para el desarrollo de nuevos proyectos, puesto que la facilidad, rapidez y capacidad de este modelo para ajustarse es destacable, muy por encima del modelo tipo completion davinci con que se contaba previamente, y considerando que tienen programado liberar modelos aún más avanzados en el corto plazo, indican un horizonte prometedor para futuras investigaciones.

El hecho de conseguir resultados tan prometedores contando con un presupuesto y muestra particularmente acotado, demuestra que los algoritmos en desarrollo poseen el potencial para transformar radicalmente la interpretación de contratos en el sistema judicial chileno, aspirando a respuestas más pulidas, estandarizadas y, en última instancia, a un sistema más equitativo.

Finalmente, lo más interesante pasa por el hallazgo de que es posible alimentar al chatbot respecto a las decisiones que en el pasado ha ido tomando la Corte Suprema, y que aquello efectivamente influye en las respuestas que se ofrecen al usuario. Teniendo en cuenta dicha aproximación parece factible el desarrollo de un modelo de procesamiento de lenguaje natural para la interpretación judicial de contratos que satisfaga las

expectativas, tanto de los juristas, como de quienes recurren ocasionalmente a los tribunales de justicia para zanjar las controversias.

Ahora bien, estos resultados actualmente se limitan a las capacidades de los mejores modelos que existen actualmente, los cuales son los chatbot proporcionados por OpenAI, GPT-3.5 y GPT-4, mientras que los demás modelos de la competencia están más lejos de llegar a vislumbrar la misma calidad de resultados, dicho esto, no quiere decir que estén demasiado lejos de lograrlo, dado que a lo largo del último año, ha habido avances particularmente rápidos en el área, incrementando las expectativas de desarrollo en el corto plazo.

Para próximas iteraciones de proyectos, sería necesario un equipo de abogados más vasto para construir una base de datos más grande, donde se estudie a profundidad el estilo y tipo de razonamiento de la Corte Suprema, y cómo presentarle el problema de manera resumida al modelo, de tal manera de rescatar la mejor respuesta de éste e intentar mejorar sobre ello, o en su defecto, el utilizar un reducido equipo de profesional del rubro, que se dediquen a iterar sobre los modelos para encontrar la mejor manera de conseguir la mejor respuesta posible en base a los estándares esperados que se definan previamente. O incluso, se podría ver la posibilidad de realizar algún tipo de proyecto en conjunto con los miembros de la Corte Suprema, para definir los estándares esperados.

Otras consideraciones sobre las posibles limitaciones futuras de un proyecto de estas características son, en primer lugar, es posible que no tengamos un sistema judicial lo suficientemente parejo como para que el algoritmo aprenda un estilo clásico de razonamiento, sumado a que, si las leyes cambian, el razonamiento previo adquirido podría llegar a quedar obsoleto. Es por esto que, en principio, es recomendable trabajar sobre un nicho más específico y de mayor expertiz para lograr resultados más adecuados y consistentes en el tiempo.

Bibliografía

- [1] Poder Judicial - ¿Qué es el Poder Judicial? (s. f.). <https://www.pjud.cl/post/que-es-el-poder-judicial>.
- [2] Coloma, Rodrigo. Interpretación de contratos: entre literalidad e intención /en/ Revista Chilena de Derecho Privado, N°26, 2016, pp. 9 – 47
- [3] What is CRISP DM? (s. f.) <https://www.datascience-pm.com/crisp-dm-2/>
- [4] Leyes (s. f.) https://leyes-cl.com/codigo_civil.html
- [5] Poria, S., Hazarika, D., Majumder, N., & Mihalcea, R. (2020). Beneath the tip of the iceberg: Current challenges and new directions in sentiment analysis research. IEEE Transactions on Affective Computing.
- [6] Jurafsky, D., & Martin, J. H. (2009). Speech and Language Processing (2nd ed.). Prentice Hall.
- [7] Salton, G., & McGill, M. J. (1986). Introduction to modern information retrieval. McGraw-Hill, Inc.
- [8] Jatav, V., Teja, R., Bharadwaj, S., & Srinivasan, V. (2017). Improving part-of-speech tagging for NLP pipelines. arXiv preprint arXiv:1708.00241.
- [9] Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., & Dyer, C. (2016). Neural architectures for named entity recognition. arXiv preprint arXiv:1603.01360.
- [10] Salton, G., Wong, A., & Yang, C. (1975). A vector space model for automatic indexing. Commun. ACM, 18, 613-620.
- [11] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. Advances in neural information processing systems.
- [12] Le, Q., & Mikolov, T. (2014, June). Distributed representations of sentences and documents. In International conference on machine learning (pp. 1188-1196). PMLR.
- [13] Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis.
- [14] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation.
- [15] Andor, D., Alberti, C., Weiss, D., Severyn, A., Presta, A., Ganchev, K., ... & Collins, M. (2016). Globally normalized transition-based neural networks. arXiv preprint arXiv:1603.06042
- [16] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

- [17] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
- [18] Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving Language Understanding by Generative Pre-training.
- [19] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in neural information processing systems, 33, 1877-1901.
- [20] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. The Journal of Machine Learning Research, 21(1), 5485-5551.
- [21] OpenAI (2023). GPT-4 Technical Report. ArXiv, abs/2303.08774.
- [22] Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., ... & Lample, G. (2023). Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971.
- [23] Stanford University. Alpaca: A Strong, Replicable Instruction-Following Model (s. f.) <https://crfm.stanford.edu/2023/03/13/alpaca.html>
- [24] Lmsys Org (2023). Vicuna: An Open-Source Chatbot Impressing GPT-4 with 90%* ChatGPT Quality <https://lmsys.org/blog/2023-03-30-vicuna/>
- [25] Hugging Face, Learn (s. f.) <https://huggingface.co/learn>
- [26] Lmsys Org. (s. f.) <https://lmsys.org/about/>
- [27] Chatbot Arena (s. f.) <https://chat.lmsys.org/>
- [28] Fuentes González, A. (2021). EL PROCESAMIENTO DE TEXTOS JURÍDICOS: NATURAL LANGUAGE PROCESSING & LEGAL TECH.
- [29] Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., & Androutsopoulos, I. (2020). LEGAL-BERT: The muppets straight out of law school. arXiv preprint arXiv:2010.02559.
- [30] Polo, F. M., Mendonça, G. C. F., Parreira, K. C. J., Gianvechio, L., Cordeiro, P., Ferreira, J. B., ... & Vicente, R. (2021). LegalNLP--Natural Language Processing methods for the Brazilian Legal Language. arXiv preprint arXiv:2110.15709.
- [31] Licari, D., & Comandè, G. (2022). ITALIAN-LEGAL-BERT: A Pre-trained Transformer Language Model for Italian Law. In CEUR Workshop Proceedings (Ed.), The Knowledge Management for Law Workshop (KM4LAW).
- [32] Paul, S., Mandal, A., Goyal, P., & Ghosh, S. (2022). Pre-training transformers on indian legal text. arXiv preprint arXiv:2209.06049.
- [33] Chriqui, A., Yahav, I., & Bar-Siman-Tov, I. (2022). Legal hebert: A bert-based nlp model for hebrew legal, judicial and legislative texts. Judicial and Legislative Texts (June 27, 2022).

- [34] Case Text, CoCounsel (s. f.) <https://casetext.com/cocounsel/>
- [35] Tirant, Sofia (s. f.) <https://sofiachile.tirant.com/>
- [36] Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., ... & Scialom, T. (2023). Llama 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv:2307.09288.
- [37] Penedo, G., Malartic, Q., Hesslow, D., Cojocaru, R., Cappelli, A., Alobeidli, H., ... & Launay, J. (2023). The RefinedWeb dataset for Falcon LLM: outperforming curated corpora with web data, and web data only. arXiv preprint arXiv:2306.01116.
- [38] h2oGPT (s. f.) <https://gpt.h2o.ai/>
- [39] OCR PDF (s. f.) <https://www.ilovepdf.com/es/ocr-pdf>
- [40] FreedomGPT (s. f.) <https://www.freedomgpt.com/>
- [41] Text generation web UI (s. f.) <https://github.com/oobabooga/text-generation-webui>
- [42] OpenAI, Fine-tuning (s. f.) <https://platform.openai.com/docs/guides/fine-tuning>