



UNIVERSIDAD TÉCNICA DEL NORTE
FACULTAD DE INGENIERÍA EN CIENCIAS
APLICADAS
CARRERA DE TECNOLOGÍAS DE LA
INFORMACIÓN

TRABAJO DE INTEGRACIÓN CURRICULAR

TEMA:

DESARROLLO DE UN SISTEMA DE ANÁLISIS AUTOMÁTICO DE
TÉRMINOS Y CONDICIONES DE SOFTWARE MEDIANTE EL USO
DE MODELOS DE LENGUAJE EXTENSO (LLMs)

Trabajo de titulación previo a la obtención del título en Ingeniero en
Tecnologías de la Información

Línea de investigación: Desarrollo, aplicación de software y cyber
security (seguridad cibernética)

AUTOR:

Luna Córdova Leonardo Javier

DIRECTOR:

Msc. Pablo Andrés Lanteta Lopez

Ibarra – Ecuador 2025



UNIVERSIDAD TÉCNICA DEL NORTE

BIBLIOTECA UNIVERSITARIA

AUTORIZACIÓN DE USO Y PUBLICACIÓN A FAVOR DE LA UNIVERSIDAD TÉCNICA DEL NORTE

1. IDENTIFICACIÓN DE LA OBRA

En cumplimiento del Art. 144 de la Ley de Educación Superior, hago la entrega del presente trabajo a la Universidad Técnica del Norte para que sea publicado en el Repositorio Digital Institucional, para lo cual pongo a disposición la siguiente información:

DATOS DE CONTACTO			
CÉDULA DE IDENTIDAD:	0401843263		
APELLIDOS Y NOMBRES:	Luna Córdova Leonardo Javier		
DIRECCIÓN:	Av. Jesús Yerovi 227 y Quito		
EMAIL:	ljlunac@utn.edu.ec		
TELÉFONO FIJO:		TELÉFONO MÓVIL:	0982171945

DATOS DE LA OBRA	
TÍTULO:	DESARROLLO DE UN SISTEMA DE ANÁLISIS AUTOMÁTICO DE TÉRMINOS Y CONDICIONES DE SOFTWARE MEDIANTE EL USO DE MODELOS DE LENGUAJE EXTENSO (LLMs)
AUTOR (ES):	Luna Córdova Leonardo Javier
FECHA: DD/MM/AAAA	04/09/2025
SOLO PARA TRABAJOS DE GRADO	
PROGRAMA:	<input checked="" type="checkbox"/> PREGRADO <input type="checkbox"/> POSGRADO
TÍTULO POR EL QUE OPTA:	Ingeniero en Tecnologías de la Información
ASESOR /DIRECTOR:	Phd. Cathy Guevara Vega/ Msc. Pablo Andrés Landeta

2. CONSTANCIAS

El autor (es) manifiesta (n) que la obra objeto de la presente autorización es original y se la desarrolló, sin violar derechos de autor de terceros, por lo tanto, la obra es original y que es (son) el (los) titular (es) de los derechos patrimoniales, por lo que asume (n) la responsabilidad sobre el contenido de la misma y saldrá (n) en defensa de la Universidad en caso de reclamación por parte de terceros.

Ibarra, a los 08 días del mes de septiembre de 2025

EL AUTOR:

(Firma).....

Nombre: Luna Córdova Leonardo Javier

CERTIFICACIÓN DEL DIRECTOR DEL TRABAJO DE INTEGRACIÓN CURRICULAR

Ibarra, 04 de septiembre de 2025.

MSc. Pablo Andrés Landeta
DIRECTOR DEL TRABAJO DE INTEGRACIÓN CURRICULAR

CERTIFICA:

Haber revisado el presente informe final del trabajo de Integración Curricular, el mismo que se ajusta a las normas vigentes de la Universidad Técnica del Norte; en consecuencia, autorizo su presentación para los fines legales pertinentes.



MSc. Pablo Andrés Landeta

C.C: 1002161055

APROBACIÓN DEL COMITÉ CALIFICADOR

El Comité Calificado del trabajo de Integración Curricular “ Desarrollo de un sistema de análisis automático de términos y condiciones de software mediante el uso de Modelos de Lenguaje Extenso (LLMs)” elaborado por Leonardo Javier Luna Córdova , previo a la obtención del título del Ingeniero en Tecnologías de la Información , aprueba el presente informe de investigación en nombre de la Universidad Técnica del Norte:



MSc. Pablo Andrés Landeta

C.C: 1002161055



Dra. Cathy Pamela Guevara

C.C: 1002334835

DEDICATORIA

A mis padres, quienes confiaron en mi capacidad para alcanzar cualquier meta que me propusiera, sin importar qué tan complejas parecieran su apoyo incondicional y las enseñanzas de perseverancia fueron el motor que me impulsó cuando los desafíos técnicos parecían imposibles de resolver me enseñaron que los obstáculos son oportunidades disfrazadas y que la dedicación siempre encuentra su recompensa.

A mis hermanos, que celebraron cada pequeño avance en mi investigación como si fuera suyo propio sus palabras de aliento llegaron justo cuando más las necesitaba, recordándome que nunca estoy solo en esta travesía. Con ustedes aprendí que el éxito tiene más sabor cuando se comparte en familia.

A mi novia, quien entendió mis largas horas frente a la pantalla sin quejas, apoyándose con paciencia infinita durante todo este proceso académico. Tu comprensión, cariño en las madrugadas fueron fundamentales para que pudiera concentrarme en desarrollar este proyecto. Gracias por creer en mis sueños y por hacer que cada desafío se convirtiera en una oportunidad de crecimiento conjunto.

A mis amigos de la universidad, quienes siempre estuvieron ahí para escuchar mis frustraciones y celebrar conmigo cada vez que algo funcionaba correctamente. Con ustedes descubrí que el conocimiento compartido se multiplica y que las mejores soluciones nacen del trabajo en equipo y la amistad sincera.

A todos ustedes, que creyeron en mí incluso cuando yo mismo dudaba si podría lograrlo. Este trabajo no es solo mío, porque sin su apoyo, estas páginas jamás habrían cobrado vida. En cada línea de esta investigación está reflejado el amor y la confianza que me brindaron.

Luna Córdova Leonardo Javier

AGRADECIMIENTO

Esta etapa académica llega a su fin. Quiero expresar mi eterno agradecimiento.

Al MSc. Pablo Andrés Landeta, director de este trabajo y a la Dra. Cathy Guevara asesora, les agradezco su orientación y paciencia durante la investigación. Su experiencia me ayudó a dar forma a este proyecto. Sus conocimientos especializados fueron clave para mi crecimiento profesional.

Mi familia me ha dado amor incondicional y comprensión durante estos años de estudio. Su fe en mis capacidades me impulsó a seguir adelante. En los momentos más difíciles, su aliento fue mi motor para alcanzar este logro académico.

Luna Córdova Leonardo Javier

ÍNDICE DE CONTENIDOS

DEDICATORIA	II
AGRADECIMIENTO	III
ÍNDICE DE FIGURAS	IX
ÍNDICE DE TABLAS	IX
RESUMEN	XI
ABSTRACT	XII
CAPÍTULO I	
INTRODUCCIÓN	18
1.1 Planteamiento del Problema	18
1.2 Objetivos	19
1.2.1 Objetivo General	19
1.2.2 Objetivos Específicos	19
1.3 Alcance y delimitación	19
1.4 Justificación	20
CAPÍTULO II	
MARCO TEÓRICO	23
2.1 Antecedentes	23
2.2 Bases teóricas	24
2.2.1 Modelo de Lenguaje Grande y Procesamiento de Lenguaje Natural (NLP).	24
2.2.2 Modelo de Lenguaje (LLMs)	24
2.2.3 Procesamiento de Lenguaje Natural (NLP)	26
2.2.4 Evaluación de Modelos de Lenguaje	27
2.2.5 Pre-entrenamiento y Ajuste Fino.	28
2.2.6 Generación de Texto	28
2.2.7 Innovaciones y Limitaciones en el Procesamiento de Lenguaje Natural	29

2.2.8	Modelos de Lenguaje y su Impacto en la IA.	29
2.2.9	LM Studio en el Procesamiento de Lenguaje Natural y su Aplicación en Términos y Condiciones de Software.	30
2.2.10	Aspectos Legales y Éticos.	31
2.2.11	Privacidad y Autonomía Personal.	31
2.2.12	Consentimiento Informado	32
2.2.13	Seguridad de la Información.	33
2.2.14	Derechos de los Titulares de Datos.	34
2.2.15	Bases Legales para el Tratamiento de Datos.	35
2.2.16	Ética en la Inteligencia Artificial	36
2.2.17	Perspectivas Tecnológicas y Sociales	37
2.2.18	Teoría de la Interacción Humano-Computadora (HCI).	37
2.2.19	Innovación Disruptiva	38
2.2.20	Teoría de la Evaluación de Originalidad.	38
2.2.21	Modelos de Precios en Servicios en la Nube	39
2.2.22	Evaluación Empírica en Ingeniería de Software.	39
2.2.23	Confianza en la Tecnología.	40
 CAPÍTULO III		
	MATERIALES Y MÉTODOS	41
3.1	Tipo de Investigación	41
3.1.1	Metodología	41
3.1.2	Varibales de Estudio	42
3.2	Población y Muestra	43

3.2.1	Población	43
3.2.2	Muestra	44
3.3	Diseño de la Investigación	45
3.3.1	Materiales, Equipos y Software	45
3.3.2	Entorno de desarrollo del sistema	46
3.4	Metodología de desarrollo	48
3.4.1	Cronograma de Desarrollo	48
3.4.2	Fase de Análisis y Definición de Requisitos	52
3.4.3	Fase de diseño del sistema	53
3.4.4	Diseño detallado de la estructura de directorios	55
3.4.5	Requisitos Funcionales	55
3.4.6	Requisitos No Funcionales	57
3.4.7	Fase de implementación	58
3.4.8	Fase de verificación y pruebas	62
3.4.9	Fase de Despliegue y Mantenimiento	77
 CAPÍTULO IV		
RESULTADOS Y ANÁLISIS		
4.1	Resultados de la evaluación con G-Eval	79
4.1.1	Desarrollo y Adaptación del Cuestionario de Evaluación	79
4.1.2	Métricas principales de evaluación	79
4.1.3	Correlación entre métricas	81
4.1.4	Rendimiento por proveedor de servicio	83
4.2	Análisis Detallado de Rendimiento del Sistema	84

4.2.1	Estadísticas de tiempos de procesamiento	84
4.2.2	Rendimiento por tamaño de documento	86
4.2.3	Eficiencia del sistema de caché.	87
4.2.4	Documentos con mayor mejora de rendimiento	88
4.2.5	Comparativa de documentos	88
4.2.6	Evolución de patrones en consultas consecutivas.	89
4.2.7	Patrones de evolución en consultas sucesivas.	90
4.2.8	Impacto del hardware en el rendimiento	90
4.3	Correlación entre Métricas de Rendimiento y Calidad de Análisis	90
4.3.1	Tiempos de procesamiento vs. precisión del análisis	91
4.3.2	Efecto del tamaño del documento en la calidad del análisis.	91
4.3.3	Impacto del sistema de caché en la calidad de las respuestas.	92
4.3.4	Rendimiento vs. calidad por proveedor.	92
	CONCLUSIONES	94
	RECOMENDACIONES	95
	BIBLIOGRAFÍA	97
	ANEXOS	102

ÍNDICE DE FIGURAS

Figura 1	Modelos de lenguaje grandes. [14]	26
Figura 2	Procesamiento del lenguaje natural. [18]	27
Figura 3	Diagrama de Procesos.	43
Figura 4	Modelo de Cascada Adaptado para el Desarrollo del Sistema de Análisis Automatizado de Términos y Condiciones mediante LLMs.	48
Figura 5	Estructura de Directorios: Análisis Automático con LLMs	55
Figura 6	Procesador de documentos.	59
Figura 7	Implementación de LM Studio.	60
Figura 8	Interfaz de usuario del sistema.	62
Figura 9	Ejemplo de salida JSON con respuesta del modelo a una consulta específica sobre términos y condiciones.	62
Figura 10	Informe de evaluación G-EVAL simplificada: análisis de precisión, relevancia y fidelidad del resumen generado por el modelo.	72
Figura 11	Comparación de métricas de evaluación según proveedor de servicio.	76
Figura 12	Métricas de Evaluación GEval	80
Figura 13	Correlación entre Métricas y Puntuación Global	82
Figura 14	Tiempos de Procesamiento por Tipo de Operación	84
Figura 15	Curva de optimización del sistema mostrando el área de mejora progresiva. La función de eficiencia demuestra una relación exponencial inversa entre el uso del sistema de caché y los tiempos de procesamiento.	86
Figura 16	Arquitectura General del Sistema de Análisis de Términos y Condiciones	102
Figura 17	Estructura de Directorios del Proyecto	103
Figura 18	Inicialización del Sistema en Entorno Local	104
Figura 19	Interfaz de Usuario: Pantalla Principal	104
Figura 20	Selección de Documentos para Análisis	105
Figura 21	Interfaz de Carga de Documentos del Sistema	105
Figura 22	Visualización del Resumen de Términos y Condiciones	106
Figura 23	Versión Simplificada de Términos y Condiciones	107
Figura 24	Conversación sobre Términos de Licencia	108
Figura 25	Interfaz de LM Studio con Modelo LLama-3.2-3b-instruct	108
Figura 26	Consulta sobre Ingeniería Inversa en Términos Legales	109
Figura 27	Sistema de Caché del Analizador de Términos y Condiciones	109
Figura 28	Archivo de Caché JSON del Sistema	110
Figura 29	Script Evaluador de Análisis de Términos y Condiciones	110
Figura 30	Selección de Archivo de Caché en Evaluador GEval	111
Figura 31	Resultados de Evaluación GEval para Documentos Legales	111
Figura 32	Informe Detallado de Evaluación GEval para Análisis Legal	112
Figura 33	Archivo CSV de Evaluaciones GEval para Análisis de Términos y Condiciones	112
Figura 34	Ventana de Información del Sistema de Análisis	113

ÍNDICE DE TABLAS

I	Recursos y herramientas utilizados	46
II	Cronograma de actividades - fase 1	49
III	Cronograma de actividades - fase 2	49
IV	Cronograma de actividades - fase 3	50
V	Cronograma de actividades - fase 4	51
VI	Cronograma de actividades - fase 5	51
VII	RF001 carga de documentos	55
VIII	RF002 análisis automático	56
IX	RF003 interfaz de chat	56
X	RF004 visualización de resultados	56
XI	RF005 gestión de sesiones	56
XII	RNF-USA-001 diseño responsivo	57
XIII	RNF-REN-001 rendimiento	57
XIV	RNF-SEG-001 seguridad de datos	57
XV	Clasificación de modelos de lenguaje por categoría de tamaño	63
XVI	Arquitecturas populares de LLMs	64
XVII	Impacto del fine-tuning en instrucciones	64
XVIII	Evaluación de capacidades específicas de LLMs	65
XIX	Modelos seleccionados para la evaluación	65
XX	Precisión en identificación de cláusulas clave (0-100 %)	66
XXI	Tiempo de procesamiento promedio (segundos por documento)	66
XXII	Requisitos computacionales por modelo	67
XXIII	Técnicas de eficiencia para modelos de lenguaje según Zhao	67
XXIV	Trade-offs en la selección de modelos según Zhao	68
XXV	Métricas de rendimiento del sistema evaluadas en documentos de términos y condiciones de servicios tecnológicos	75
XXVI	Resultados de métricas de evaluación mediante G-EVAL	80
XXVII	Coefficientes de correlación entre métricas y puntuación global	81
XXVIII	Comparativa de métricas por proveedor	83
XXIX	Tiempos de ejecución por tipo de operación	85
XXX	Procesamiento por tamaño de documento	87
XXXI	Métricas de rendimiento del sistema	87
XXXII	Comparativa de tiempos entre primera y última consulta por documento	88
XXXIII	Documentos con tiempos de procesamiento extremos	89
XXXIV	Puntuación global promedio por categoría de tamaño	91
XXXV	Comparativa de métricas con y sin uso de caché	92
XXXVI	Tiempo de procesamiento y puntuación global por proveedor	92

RESUMEN

Contexto: Las aplicaciones SaaS tienen Acuerdos de Cliente (CA) que son complicados de entender para los usuarios. De hecho, los usuarios no están motivados a leer los CA, aunque es importante conocer cuáles son sus obligaciones, derechos y prohibiciones en relación con la aplicación SaaS. **Objetivo:** Desarrollar un sistema automatizado basado en Modelos de Lenguaje Grandes (LLMs) para analizar y extraer cláusulas clave de los términos y condiciones del software, permitiendo a los usuarios entender rápidamente sus derechos. **Método:** Se desarrolló una aplicación avanzada basada en LLM utilizando el modelo llama-3.2-3b-instruct, implementado con FastAPI y LM Studio para la ejecución local del modelo, garantizando la privacidad de datos y un procesamiento eficiente. Para evaluar los resultados, se analizaron 19 documentos, generando 48 respuestas individuales. Se emplearon Evaluadores Basados en Modelos para evaluar el rendimiento del sistema en múltiples dimensiones críticas utilizando el framework G-Eval. **Resultados:** El sistema de análisis automatizado demostró un rendimiento notable: Precisión de Contexto 84,68 %, Relevancia de Respuesta 88,62 %, Fidelidad del Documento 75,64 %, Recuperación de Contexto 64,57 % y Puntuación de Rendimiento Global 79,21 %. Estas métricas revelan la capacidad del sistema para transformar documentos legales complejos en información comprensible, cerrando la brecha entre el lenguaje legal técnico y la comprensión del usuario. **Conclusiones:** Los resultados validan el potencial de los modelos de lenguaje avanzados para simplificar los términos y condiciones del software. Con alta relevancia de respuesta (86,77 %) y precisión de contexto (82,92 %), el sistema muestra un potencial significativo para democratizar el acceso a la información legal.

Palabras clave: Modelo de Lenguaje Grande, Análisis Automatizado, Acuerdos de Cliente, Procesamiento de Lenguaje Natural, G-Eval

ABSTRACT

Context: SaaS applications have Customer Agreements (CA) that are complicated for users to understand. As a matter of fact, users are unmotivated to read the CAs, even though it is important to know what their obligations, rights, and prohibitions are in relation to the SaaS application. **Aim:** Develop an automated system based on Large Language Models (LLMs) to analyze and extract key clauses from software terms and conditions, enabling users to quickly understand their rights. **Method:** An advanced LLM-based application was developed using the llama-3.2-3b-instruct model, implemented with FastAPI and LM Studio for local model execution, ensuring data privacy and efficient processing. To evaluate the results, 19 documents were analyzed, generating 48 individual responses. Model-Based Scorers were employed to evaluate the system's performance across multiple critical dimensions using G-Eval framework. **Results:** The automated analysis system demonstrated notable performance: Context Precision 84,68 %, Response Relevance 88,62 %, Document Fidelity 75,64 %, Context Recall 64,57 % and Global Performance Score 79,21 %. These metrics reveal the system's capability to transform complex legal documents into comprehensible insights, bridging the gap between technical legal language and user understanding. **Conclusions:** The results validate the potential of advanced language models in simplifying software terms and conditions. With high response relevance (86.77 %) and context precision (82.92 %), the system shows significant potential for democratizing legal information access.

Keywords: Large Language Model, Automated Analysis, Customer Agreements, Natural Language Processing, G-Eval

CAPÍTULO I

INTRODUCCIÓN

1.1 Planteamiento del Problema

En la actualidad, millones de personas están aceptando los términos y condiciones de uso del software sin haber leído o comprendido el contenido de estas cláusulas. Todo esto sucede porque estos documentos son a menudo extensos y están redactados en un lenguaje técnico y legal que no siempre es de fácil comprensión. Por esta causa, el usuario no siempre conoce sus derechos, obligaciones, o los riesgos que está aceptando si se da a esto última una aceptación.

Tal desconocimiento puede llevar a situaciones en las cuales los usuarios se ven en situación de incumplir las reglas de servicio o, lo que es peor, aceptar condiciones que limitan su privacidad, su seguridad, sus derechos, etc. sin tenerlo en cuenta. Ver también que, en caso de conflicto, los usuarios pueden partir de la premisa que desconocen cuál sera la cláusula que protege al proveedor o en qué condiciones podra solicitar soporte o a la inversa, solicitar un reembolso en caso de que se haya realizado la compra de un servicio.

De este modo, la falta de conocimiento hace que el establecimiento de relaciones entre el usuario y el proveedor de servicio presente diversidad de problemas. En este nuevo horizonte, se necesita de un sistema automatizado que mediante el uso de modelos de lenguaje extenso (LLMs) permita analizar los acuerdos de uso y formular respuestas a interrogantes específicas. Este sistema permitirá al usuario poder con inmediatez y facilidad acceder a cuestiones sobre lo que significan determinados términos en relación con los derechos que puedan corresponderles, así como los que pueden corresponder al proveedor.

1.2 Objetivos

1.2.1 Objetivo General

Desarrollar un sistema automatizado que utilice Modelos de Lenguaje Extenso (Large Language Models, LLMs) para analizar y extraer las cláusulas clave en los términos y condiciones del software para que los usuarios comprendan los derechos, obligaciones y riesgos de manera rápida y fácil.

1.2.2 Objetivos Específicos

- Investigar el modelo de lenguaje extenso (LLM) que mejor se adapte a la identificación y análisis de cláusulas clave en términos y condiciones de software considerando opciones de LLM open-source que permitan una implementación económica y eficiente.
- Desarrollar una aplicación que utilice LLMs para realizar el análisis automático de los acuerdos de cliente (CA) extrayendo y categorizando cláusulas importantes para que sea más fácil de comprender.
- Evaluar la aplicación utilizando (Model Based Scorers) con el objetivo de valorar la relevancia y la precisión de las respuestas generadas por el LLM en comparación con los estándares esperados.

1.3 Alcance y delimitación

En cuanto a esta investigación, desarrollar una aplicación web que aproveche los Modelos de Lenguaje Grande (LLMs) constituye el objetivo fundamental. Ejecutar análisis automatizado sobre los Términos y Condiciones de software representa la funcionalidad principal de dicha aplicación, tal como se ha explicado previamente. Con respecto a los usuarios, estos podrán cargar documentos tanto en español como en inglés. Conseguir un análisis automatizado que identifique y resuma términos clave como derechos, obligaciones y restricciones constituye el

beneficio que recibirán, según lo mencionado anteriormente. Mejorar su comprensión de los acuerdos representa el propósito final de este proceso, de acuerdo con lo expuesto.

La complejidad técnica y los costos de implementación, emplear un LLM de código abierto sin modificaciones ni entrenamiento adicional constituye la estrategia para compensar estos aspectos, tal como se ha explicado previamente en lo que se refiere a la aplicación web mencionada anteriormente, esta no posibilita características avanzadas como almacenamiento de documentos o comparación de múltiples acuerdos. Simplificar la interacción y comprensión del usuario representa el propósito de esta decisión, según lo mencionado, de acuerdo con lo expuesto.

La validación del proyecto se ejecutará mediante pruebas básicas de precisión y relevancia. Conseguir que el sistema logre su objetivo de extraer la información más relevante de los documentos constituye el propósito fundamental. En cuanto al sistema, este se concibió como apoyo y no como sustituto de una revisión legal. Por esta razón, no se contemplan evaluaciones exhaustivas o validaciones legales formales. Con respecto a la puesta en marcha, esta tendrá lugar en un servidor web que suministre acceso amigable y funcional para los usuarios. El proyecto aceptará únicamente documentos en inglés o español. No obstante, adaptarse a otros idiomas en el futuro constituye una posibilidad.

El sistema se centrará exclusivamente en términos y condiciones correspondientes a software comercial. Aplicarse a otros tipos de documentos legales o acuerdos sin relación directa con software no constituye parte del alcance de esta investigación.

La herramienta estará diseñada para usuarios con desconocimiento legal, por lo que se hará una interpretación sencilla y entendible de los términos y condiciones. No estará dirigida a especialistas legales o abogados [1]-[3].

1.4 Justificación

La propuesta de crear un sistema de análisis automático de términos y condiciones de software se basa de forma persuasiva en la implementación de Modelos de Lenguaje Extenso (LLMs). Las razones que motivan este estudio se deben a la creciente complejidad de los contratos de

software, que frecuentemente son amplios y complicados de entender para el usuario común. La ausencia de transparencia en estos documentos perjudica el consentimiento informado y amenaza la privacidad y seguridad de los usuarios, quienes pueden aceptar términos sin comprender sus consecuencias legales.

Esta propuesta se encuentra alineada con los principios y objetivos del Plan Nacional de Telecomunicaciones y Tecnologías de Información del Ecuador (2016-2021), que enfatiza la democratización de las tecnologías de la información y la promoción del acceso universal a las mismas. Dentro del marco del plan nacional, encontramos un resultado particularmente relevante para nuestra investigación: asegurar que los medios técnicos de información y comunicación sirvan como catalizadores efectivos para el desarrollo tanto del crecimiento social como económico en el país.[4].

Una función central de la herramienta elaborada consiste en reducir las brechas informacionales que históricamente han separado a los proveedores de software de sus usuarios finales. Alinearse con los principios de la política de inclusión digital constituye una característica notable del sistema, pues dicha política busca que la tecnología sea accesible y comprensible para diversos sectores poblacionales. Adicionalmente, el Plan Nacional enfatiza el uso de las TIC como instrumentos para optimizar servicios públicos y privados. La aplicación diseñada posee el potencial de establecer relaciones más equilibradas entre las partes que participan en acuerdos de software. [4].

En esta investigación se creó un sistema que analiza automáticamente los términos y condiciones de software. Los resultados que arroja son mucho más claros y fáciles de entender que los documentos originales. Con esto, los usuarios pueden entender mejor tres cosas clave: qué derechos tienen, qué responsabilidades asumen y qué riesgos corren. Las empresas también ganan porque pueden ofrecer más transparencia a sus usuarios, lo que mejora su experiencia.[5].

Abordar uno de los problemas más persistentes en el ámbito digital constituye el objetivo central de este sistema: la falta de claridad que caracteriza a muchos términos y condiciones de software. Al hacerlo, buscamos fomentar un acceso notablemente más equitativo y, sobre

todo, más comprensible a esa información contractual que tantas veces resulta inaccesible para el usuario común. Desde el punto de vista técnico, consideramos que este proyecto puede aportar beneficios significativos al campo de la ingeniería, particularmente a través de nuestra exploración de las capacidades que poseen los LLMs para el análisis de textos jurídicos[6].

Establecemos un sistema que opera bajo estrictos principios éticos y jurídicos, manteniéndose en plena concordancia con las regulaciones de privacidad vigentes en nuestro país, específicamente la Ley Orgánica de Protección de Datos Personales (LOPDP) de Ecuador.

Ahora bien, uno de los aspectos que consideramos más valiosos de nuestro sistema radica en su potencial para transformar esa relación frecuentemente problemática entre usuarios y proveedores de servicios digitales de este modo el sistema actúa como herramienta de apoyo esencial para usuarios que necesitan comprender qué están aceptando.

La mayoría terminamos haciendo clic en **acepto** sin conocer las implicaciones reales. Situación preocupante cuando cada aceptación ciega compromete aspectos críticos de la privacidad personal, derechos de consumidores, seguridad financiera de ahí que crear una solución que democratice el acceso a información crítica aporta un avance técnico notable, mientras contribuye genuinamente al bienestar de usuarios digitales. La herramienta responde a una necesidad concreta que permitir que personas sin formación jurídica especializada comprendan documentos legales complejos y tomen decisiones informadas sobre servicios digitales de uso cotidiano.

Los avances metodológicos obtenidos facilitan nuevas investigaciones futuras en este sentido diversos campos necesitan interpretar documentos técnicos especializados de forma automática y pueden aprovechar estos resultados, donde nuestro trabajo funciona como base sólida para desarrollos posteriores. Estos precedentes aportan valor considerable a aplicaciones que exigen alta precisión y confiabilidad[7], [8]en efecto la meta trasciende el simple procesamiento textual busca alcanzar exactitud en contextos profesionales y académicos.

CAPÍTULO II

MARCO TEÓRICO

2.1 Antecedentes

Durante los años 80, los convenios EULA comenzaron a integrarse en productos de software. Tales documentos destacan por su vocabulario técnico y amplio alcance, mientras los usuarios enfrentan dificultades para entender sus derechos y responsabilidades. La expansión de plataformas digitales y servicios en línea ocurrió a lo largo de los años 2000. Los términos y condiciones se volvieron más complejos abordando privacidad de información, acceso a funciones, restricciones de responsabilidad. Este tipo de documento representa una práctica que amenaza derechos de usuarios. Sin comprender totalmente las consecuencias, podrían aceptar situaciones perjudiciales para privacidad y seguridad. La evolución histórica evidencia creciente complejidad. Se requieren herramientas especializadas para facilitar comprensión por parte de usuarios no especializados en aspectos jurídicos [9].

Durante los primeros años de 2010, las organizaciones recolectaban datos personales de manera masiva. Los servicios en línea se expandían rápidamente. La comprensión pública sobre términos y condiciones era mínima múltiples escándalos de privacidad lo confirmaron. Cambridge Analytica representa el ejemplo más significativo. En 2018 se reveló que utilizaron información personal de millones de usuarios de Facebook para influir en elecciones políticas. Todo esto sin consentimiento. Este escándalo provocó despertar mundial. Los derechos de privacidad cobraron relevancia. También se evidenciaron los riesgos de aceptar términos sin comprenderlos completamente [10]. Las consecuencias fueron inmediatas. Los usuarios comenzaron a cuestionar las prácticas de recolección de datos. Las empresas enfrentaron presión regulatoria sin precedentes.

2018 marcó un punto de inflexión. La Unión Europea lanzó el GDPR respuesta directa a los problemas de privacidad digital que se acumulaban. El Reglamento General de Protección de Datos exige transparencia. Las organizaciones deben explicar claramente cómo recolectan datos. También cómo los usan. Los usuarios ganaron control sobre su información personal. Las multas por incumplimiento son severas — esto cambió las reglas del juego. Pero el GDPR hizo más que regular Europa. Otras jurisdicciones siguieron el ejemplo. Legislaciones similares aparecieron mundialmente. El panorama de protección de datos se transformó por completo [11].

2020 trajo las tecnologías de IA y LLMs. Respondían a una necesidad clara: procesar documentos complejos de forma automática. La inteligencia artificial había crecido enormemente durante la década anterior. GPT-3 y GPT-4 mostraron capacidades extraordinarias. Podían analizar términos y condiciones complejos. Los simplificaban automáticamente. Usuarios sin formación legal podían entender documentos que antes resultaban incomprensibles. Este potencial representó un avance crucial en la democratización del acceso a información legal.

Estos modelos detectan puntos clave en documentos extensos y los resumen de forma eficiente. Tal capacidad responde a una demanda concreta: la necesidad de transparencia en términos y condiciones. Las capacidades tecnológicas avanzadas se encontraron con demandas sociales específicas, donde los usuarios buscaban acceso sencillo a información jurídica compleja de este modo este encuentro generó las bases ideales para el surgimiento de herramientas especializadas que democratizan el acceso legal [8], [12], [13].

2.2 Bases teóricas

2.2.1 Modelo de Lenguaje Grande y Procesamiento de Lenguaje Natural (NLP)

2.2.2 Modelo de Lenguaje (LLMs)

Los LLM constituyen un avance significativo en inteligencia artificial, particularmente en NLP específicamente emplean redes neuronales profundas tanto para comprender como generar texto fundamentalmente el aprendizaje no supervisado establece la base metodológica, donde los sistemas se entrenan con volúmenes masivos de datos aspecto fundamental del proceso y

durante este entrenamiento obtienen conocimiento sobre patrones complejos del lenguaje humano. Las conexiones semánticas se vuelven evidentes para estos modelos. Su diseño permite captar sutilezas que antes eran imposibles de procesar automáticamente.

Predecir la siguiente palabra en una secuencia representa el núcleo operativo. El modelo absorbe gramática, contexto y significado a través de este mecanismo y funciona en múltiples escenarios y tiene capacidad predictiva sirve como cimiento. Las herramientas especializadas se construyen sobre esta base especialmente para examinar documentación técnica compleja y estos mismos documentos difíciles de interpretar se vuelven accesibles gracias a este fundamento estructural.

GPT-3 y Llama demostraron algo importante y es que estos modelos abordan una gama amplia de tareas como la desde traducción automática hasta creación de contenido creativo. Extraer conocimiento generalizable de datos de entrenamiento marca una diferencia clave y esta capacidad los distingue con el panorama de inteligencia artificial cambió profundamente. Las aplicaciones prácticas surgieron en múltiples sectores servicio al cliente, educación, producción de contenido digital. La versatilidad resultó ser su mayor fortaleza.

Estos modelos traen complicaciones serias ya que pueden generar contenido sesgado o inexacto esto representa un obstáculo principal como adoptar metodologías éticas, durante la implementación se vuelve crucial. Los desafíos metodológicos requieren validación estricta como la fiabilidad de resultados debe asegurarse las aplicaciones donde la exactitud es decisiva necesitan atención particular. [13].

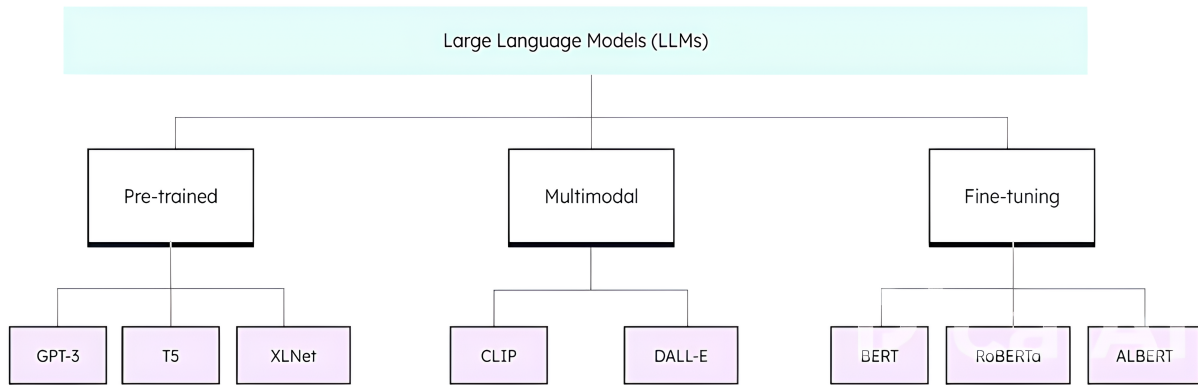


Fig. 1.

Modelos de lenguaje grandes. [14]

2.2.3 Procesamiento de Lenguaje Natural (NLP)

NLP combina lingüística, inteligencia artificial e informática. Las máquinas pueden comprender, interpretar y generar lenguaje humano efectivamente gracias a este campo interdisciplinario. Análisis automatizado de contratos encontró en NLP una herramienta clave. Extrae información y clasifica cláusulas. Los textos legales se dividen en componentes significativos esto ayuda a identificar disposiciones, obligaciones y derechos importantes.

Transformer cambió todo BERT y T5 revolucionaron NLP [15], [16] ya que proporcionan comprensión profunda del contexto y significado de oraciones. Estos modelos procesan grandes volúmenes de texto al mismo tiempo aprenden patrones complejos y en los documentos legales requieren precisión y comprensión contextual aquí brillan especialmente.

Estos modelos ejecutan clasificación de texto, detección de entidades y generación de lenguaje natural y su adopción se extendió a múltiples aplicaciones incluyendo automatización del análisis de contratos. Las capacidades diversas impulsaron su uso generalizado en el procesamiento de documentos legales.

NLP en el ámbito jurídico presenta desafíos específicos. La ambigüedad del lenguaje legal complica las cosas también las diferencias en redacción de contratos y el conocimiento especializado necesario para interpretar terminología correctamente. Estos obstáculos dificultan la

implementación de soluciones automatizadas. Los avances continuos en tecnologías de procesamiento de lenguaje natural cambian el panorama. Modelos más sofisticados prometen mayor eficiencia y precisión en análisis de contratos y la toma de decisiones informadas en el campo legal se facilita, optimizando procesos que tradicionalmente requieren revisión manual exhaustiva [17].

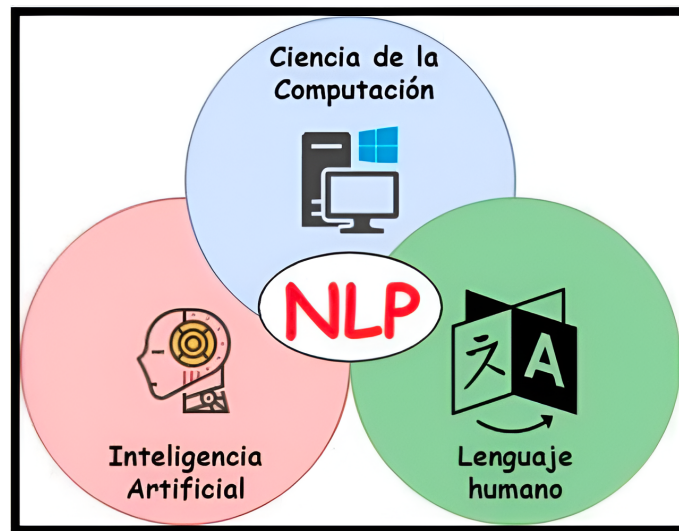


Fig. 2.
Procesamiento del lenguaje natural. [18]

2.2.4 Evaluación de Modelos de Lenguaje

Evaluar LLMs resulta crucial para entender su rendimiento y aplicabilidad real. Benchmarks estandarizados miden habilidades específicas razonamiento y comprensión del contexto principalmente. Los investigadores dependen de estas métricas para validar capacidades de los modelos.

Los modelos se clasifican según su tamaño y arquitectura. Esta categorización incluye tres tipos principales: modelos base, versiones ajustadas para instrucciones específicas y modelos conversacionales. Dicha clasificación permite a investigadores y desarrolladores reconocer capacidades particulares de cada modelo para tareas concretas. Los estudios no se limitan únicamente a medir precisión también analizan el manejo de escenarios complejos y la coherencia en

las respuestas generadas. Los resultados demuestran algo predecible: modelos de mayor tamaño exhiben rendimiento superior. Sin embargo, esta ventaja tiene un costo. Requieren infraestructura computacional más robusta y se vuelven menos accesibles para usuarios con recursos limitados.

Esta tensión define el panorama actual del campo. Por un lado, la búsqueda de capacidades avanzadas; por otro, la necesidad de democratizar el acceso. La evaluación sistemática de estos sistemas no puede detenerse. Representa un proceso fundamental para refinar tanto el diseño como la implementación práctica en sectores industriales variados. Solo mediante esta validación continua se garantiza la optimización progresiva de sus capacidades operativas[13].

2.2.5 Pre-entrenamiento y Ajuste Fino

Pre-entrenamiento y ajuste fino forman la base de los LLM ya que son grandes conjuntos de datos sin etiquetar entrenan estos modelos inicialmente aprenden representaciones universales del lenguaje durante este proceso. Después se adaptan a tareas específicas, el refinamiento utiliza datos etiquetados relevantes para cada tarea particular, este enfoque mejora el rendimiento en actividades específicas como clasificación de textos o generación de respuestas a preguntas. Su eficacia está comprobada.

La investigación muestra que el refinamiento incrementa notablemente la precisión y relevancia de respuestas del modelo, y sin embargo plantea desafíos importantes: se requieren conjuntos de datos de alta calidad, además del riesgo de sobreajuste a datos de entrenamiento. Entender estos procesos es fundamental para crear LLMs efectivos en situaciones reales, donde la aplicabilidad práctica define el éxito de cualquier implementación tecnológica [6].

2.2.6 Generación de Texto

Los LLM producen contenido coherente y relevante mediante arquitecturas avanzadas como Transformer [19], las cuales captan relaciones contextuales y semánticas entre palabras. Tales

modelos construyen oraciones gramaticalmente correctas junto con contenido informativo adecuado al contexto. La investigación demuestra que los LLM generan texto de alta calidad semejante al lenguaje humano, facilitando el desarrollo de aplicaciones para escritura automática, creación de contenido y asistencia en redacción.

Surgen problemas vinculados a la creación de contenidos sesgados o inexactos, aspecto que demanda un enfoque cuidadoso al implementar estas tecnologías en particular la generación de textos representa un área de investigación activa que busca mejorar la calidad y ética de los modelos lingüísticos en la creación de contenidos [13].

2.2.7 Innovaciones y Limitaciones en el Procesamiento de Lenguaje Natural

Los avances en técnicas de aprendizaje automático, particularmente modelos de atención y transformadores, produjeron un progreso notable en la comprensión del lenguaje dentro del NLP por ejemplo Modelos como GPT-3 y BERT demuestran efectividad en múltiples tareas de NLP al aprender representaciones del lenguaje mediante grandes volúmenes de datos no etiquetados [20], [21]. No obstante, persisten limitaciones específicas: escasez de datos para lenguajes menos representados y dificultades para interpretar el razonamiento tras las decisiones de los modelos [20], [22]no obstante las implicaciones éticas resultan fundamentales, ya que estos sistemas afectan considerablemente la sociedad y los procesos de comprensión de datos [23], [24].

2.2.8 Modelos de Lenguaje y su Impacto en la IA

Modelos como BERT y GPT-3 transformaron el procesamiento de lenguaje natural (NLP) mediante arquitecturas de redes neuronales profundas y entrenamiento con grandes volúmenes de datos, tales modelos optimizaron la comprensión y generación del lenguaje, además de exhibir capacidades emergentes que sugieren formas más generales de inteligencia artificial [20], [21], también presentan limitaciones para comprender el contexto y significado profundo, creando desafíos en tareas complejas [20], [22]. La ética en su uso resulta crucial, pues pueden perpetuar sesgos y desigualdades en sus resultados [20], [21].

2.2.9 LM Studio en el Procesamiento de Lenguaje Natural y su Aplicación en Términos y Condiciones de Software

LM Studio constituye una aplicación de escritorio que ejecuta LLMs localmente sin requerir conexión a internet y opera como interfaz especializada y proporciona control total y privacidad completa sobre los datos al evitar dependencia de servicios en la nube. La protección de datos resulta esencial en sectores corporativo y de investigación, donde mantener la confidencialidad de la información es crítico también LM Studio facilita una manera práctica y eficiente de trabajar con modelos lingüísticos mientras preserva la seguridad información-al y ofrece una experiencia personalizada [25].

Desarrollos de LM Studio:

LM Studio facilita la implementación de modelos de lenguaje en dispositivos locales, gestionando internamente datos sensibles como términos y condiciones del software y esto preserva la privacidad mientras reduce los riesgos vinculados con la gestión de datos en la nube [26]. Su manejo intuitivo permite que desarrolladores y profesionales jurídicos trabajen con LLMs sin requerir amplia comprensión técnica, simplificando la implementación de sistemas de análisis automático según las necesidades particulares del proyecto [27]. La compatibilidad con múltiples LLMs de código abierto aporta versatilidad para seleccionar el modelo más apropiado en el análisis de textos jurídicos, optimizando la precisión y relevancia de los resultados [27].

Ventajas de LM Studio:

La ejecución local de modelos de lenguaje suprime la necesidad de enviar datos a servidores externos, salvaguarda información sensible y cumple políticas estrictas de privacidad para los usuarios y logran personalizar y optimizar el rendimiento de los modelos conforme a las características de su hardware y demandas de sus proyectos, evitando las limitaciones de servicios en la nube. Respecto a accesibilidad y facilidad de uso y la interfaz gráfica de LM Studio facilita que programadores experimentados y principiantes experimenten y colaboren con LLMs, eliminando barreras técnicas considerables [25].

Características y requerimientos de LM Studio:

Requerimientos de equipo: Los LLMs requieren hardware con capacidades de procesamiento y memoria adecuadas para una implementación eficaz. Es recomendable revisar las características técnicas de cada modelo y las recomendaciones de LM Studio para garantizar un desempeño óptimo.

Administrar recursos: Al manejar modelos de gran tamaño, resulta crucial supervisar la utilización de recursos del sistema para prevenir posibles demoras o disputas con otras aplicaciones que se estén ejecutando al mismo tiempo.

LM Studio resulta útil porque:

- Garantiza privacidad total al procesar documentos legales sensibles localmente
- Elimina costos recurrentes de APIs externas para análisis de textos extensos
- Funciona sin internet, ideal para entornos con conectividad limitada
- Facilita el cumplimiento normativo al evitar transferencia de datos personales
- Permite personalización completa de parámetros del modelo para optimizar análisis legal
- Ofrece rendimiento consistente sin depender de la disponibilidad de servicios en la nube
- Se integra fácilmente con sistemas existentes mediante su API compatible con OpenAI[28].

2.2.10 Aspectos Legales y Éticos

2.2.11 Privacidad y Autonomía Personal

El enfoque de la privacidad y la autonomía personal es que todos tienen el derecho fundamental de controlar su información personal. El concepto se basa en la idea de que la privacidad es esencial para el desarrollo de la dignidad humana y la autonomía individual. En un mundo cada vez más digital donde la información se recopila, almacena y comparte a altas velocidades, la protección de datos se ha convertido en un pilar esencial para proteger el derecho a la privacidad de las personas.

La falta de protección de datos puede dar lugar a graves violaciones de la privacidad que afectan no sólo a la confianza de los ciudadanos en las instituciones, sino también a su salud emocional y mental. Las investigaciones muestran que las personas que sienten que su información personal no está segura tienen menos probabilidades de participar en actividades en línea, lo que limita su acceso a servicios y oportunidades. La divulgación de datos personales puede generar discriminación, acoso y otros daños sustanciales que afectan directamente el bienestar individual.

La Ley Orgánica de Protección de Datos Personales de Ecuador busca crear un marco legal para proteger la privacidad de los ciudadanos, iniciativa que resuelve problemas estructurales mientras promueve la confianza en el uso de servicios digitales y construye un entorno donde las personas pueden ejercer su autonomía sin temor a represalias o abusos.

Bajo tal perspectiva, la privacidad y la autonomía personal representa un elemento importante desde un punto de vista legal y también presenta implicaciones éticas y sociales que deben tenerse en cuenta en cualquier discusión sobre protección de datos [7].

2.2.12 Consentimiento Informado

El consentimiento informado constituye un principio esencial de la protección de datos donde las personas deben comprender completamente los propósitos del uso de sus datos personales y brindar consentimiento explícito para su procesamiento. Tal concepto se basa en que las personas controlen su información personal mientras toman decisiones informadas sobre su utilización. La Ley Orgánica de Protección de Datos Personales de Ecuador[29] establece que el procesamiento de datos debe fundamentarse en el consentimiento del titular.

Las organizaciones deben proporcionar información clara y comprensible sobre la recopilación, uso y distribución de los datos. Tal información incluye detalles sobre los propósitos del procesamiento, duración del almacenamiento de datos y derechos del titular sobre su información. El consentimiento informado trasciende el simple requisito legal y resulta fundamental para construir relaciones de confianza entre organizaciones e individuos, estableciendo bases sólidas para la gestión ética de la información personal.

Cuando las personas perciben control sobre su información y respeto hacia sus derechos, tienden a participar más en actividades que implican el uso de sus datos, incluyendo servicios en línea o investigaciones. Lograr un consentimiento verdaderamente informado presenta un desafío considerable, ya que las políticas de privacidad frecuentemente resultan extensas y complejas, dificultando la comprensión completa de tales documentos. Las organizaciones requieren adoptar prácticas transparentes y accesibles para garantizar que el consentimiento sea auténtico mientras los interesados se sientan empoderados respecto al procesamiento de su información personal [7].

2.2.13 Seguridad de la Información

La seguridad de la información constituye un elemento esencial para proteger datos personales, mediante el cual las organizaciones implementan medidas técnicas y organizativas que salvaguardan la información contra accesos no autorizados, asegurando confidencialidad, integridad y disponibilidad. Las ciberamenazas progresivamente más sofisticadas han revolucionado este campo, convirtiendo la seguridad de datos en una prioridad crítica dentro del entorno digital contemporáneo, particularmente para organizaciones que gestionan información personal sensible.

La Ley Orgánica de Protección de Datos Personales de Ecuador[29] establece como disposición fundamental que las organizaciones deben adoptar medidas de seguridad apropiadas para proteger los datos que procesan. Tal requisito incluye realizar una evaluación de riesgos para detectar vulnerabilidades e implementar controles de seguridad que aborden tales riesgos. Estas medidas pueden incluir el uso de cifrado, autenticación multifactor y políticas de acceso restringido, herramientas que forman parte integral de un marco de seguridad comprehensivo diseñado para salvaguardar la información sensible.

La capacitación del personal constituye un elemento esencial para garantizar que todos los empleados comprendan la importancia de la seguridad de la información y sean capaces de manejar los datos de manera responsable. Provocar errores humanos que comprometan la seguridad

de los datos representa una consecuencia directa de la falta de conocimiento de las mejores prácticas de seguridad. Las investigaciones muestran que las violaciones de seguridad constituyen una de las mayores amenazas a la protección de datos.

Las violaciones de datos pueden tener graves consecuencias para las organizaciones y las personas afectadas, incluidas pérdidas financieras, daños a la reputación y exposición de información confidencial. Las organizaciones deben tomar medidas proactivas de seguridad de la información no solo para cumplir con la ley, sino también para proteger la confianza del cliente y la integridad operativa [7], [30].

2.2.14 Derechos de los Titulares de Datos

El derecho a registrarse constituye una parte importante de la protección de datos y garantiza que las personas puedan controlar su información personal. La Ley Orgánica de Protección de Datos Personales de Ecuador[29] reconoce varios derechos de los interesados, incluido el derecho a acceder, rectificar y suprimir sus datos personales. Ser esenciales para empoderar a las personas y promover una cultura de protección de datos en la sociedad representa la función que cumplen estos derechos en el marco normativo vigente.

El derecho de acceso busca principalmente que los interesados conozcan qué datos se recopilan sobre ellos y su forma de utilización. Además de incrementar la transparencia, tal prerrogativa ayuda a las personas a detectar posibles errores o abusos durante el procesamiento de datos. El derecho de rectificación permite que los titulares corrijan información inexacta o incompleta, asegurando la exactitud y pertinencia de sus datos. También existe el derecho de supresión, conocido como derecho al olvido, donde las personas pueden solicitar la eliminación de sus datos personales bajo determinadas circunstancias específicas.

Este derecho adquiere particular relevancia cuando la información puede almacenarse en línea indefinidamente, situación donde tal persistencia impacta negativamente la reputación y vida personal de las personas. La implementación efectiva de estos derechos fortalece la confianza pública en el procesamiento de datos, además de incrementar la responsabilidad organizacional. Asegurar que los interesados puedan ejercer eficazmente sus derechos contribuye de manera

fundamental a construir un entorno donde la protección de datos sea prioritaria, beneficiando tanto a las personas como a las organizaciones que procesan sus datos [30].

2.2.15 Bases Legales para el Tratamiento de Datos

Las bases legales para el tratamiento de datos constituyen elementos fundamentales en la protección de datos, ya que establecen los fundamentos jurídicos que permiten a las organizaciones recopilar y procesar información personal. Identificar varias bases que legitiman el tratamiento de datos representa una función clave de la Ley Orgánica de Protección de Datos Personales en Ecuador, incluyendo el consentimiento del titular, el cumplimiento de obligaciones legales, y el interés legítimo del responsable del tratamiento.

El consentimiento constituye quizás la base más conocida, ya que implica que los individuos deben otorgar su autorización explícita para que sus datos sean tratados. La ley también reconoce otras bases que pueden ser igualmente importantes en diferentes contextos operativos. Permitir a las organizaciones procesar datos cuando resulta necesario para cumplir con requisitos legales o regulatorios representa la función del cumplimiento de obligaciones legales como base legitimadora del tratamiento. El interés legítimo constituye otra base que permite el tratamiento de datos, siempre que se ejecute de manera que no infrinja los derechos y libertades de los titulares. Ser especialmente relevante en contextos comerciales representa una característica de esta base, donde las organizaciones pueden necesitar procesar datos para llevar a cabo sus operaciones de manera efectiva.

Identificar y aplicar tales fundamentos jurídicos resulta fundamental para que el procesamiento de información personal se desarrolle bajo principios éticos mientras cumple estrictamente el marco normativo. Las entidades requieren mantener claridad sobre los fundamentos que sustentan sus operaciones de procesamiento de datos, permitiendo que los titulares comprendan completamente los derechos inherentes a su información personal. Tal transparencia operativa genera un ecosistema basado en confianza mutua y responsabilidad compartida para gestionar información, beneficiando tanto a las instituciones como a los individuos involucrados [7], [30].

2.2.16 Ética en la Inteligencia Artificial

La ética en IA analiza las implicaciones morales y sociales del desarrollo y uso de tecnologías como los LLM. Tales marcos éticos requieren fundamentarse en principios como honestidad, transparencia, responsabilidad y privacidad para alcanzar un progreso tecnológico responsable. La equidad opera como pilar esencial al prevenir que los sistemas de IA perpetúen sesgos o discriminación hacia grupos específicos, mientras la transparencia demanda que los usuarios comprendan el funcionamiento de estos sistemas y sus procesos de toma de decisiones. Entre los principales desafíos éticos se encuentra el riesgo de plagio junto con las consideraciones sobre derechos de propiedad intelectual que plantean estos modelos.

Los profesionales forenses legales cuentan con capacitación especializada para gestionar grandes volúmenes de datos, muchos de ellos bajo protección de derechos de autor. Esta situación genera preocupaciones importantes sobre cómo proteger los derechos de quienes crean contenido original.

Integrar especialistas con conocimientos jurídicos avanzados en ingeniería informática trae efectos contradictorios en el empleo: mejora la eficiencia operacional pero también elimina ciertos puestos de trabajo. Capacitar y apoyar a los trabajadores afectados representa una obligación ética central para la IA, garantizando así que la transición laboral sea justa y equilibrada.

Desarrollar marcos normativos y regulatorios resulta esencial para guiar el uso de estas tecnologías y establecer estándares de calidad y seguridad en el desarrollo de código. Estos marcos incluyen auditorías y evaluaciones de impacto ético que buscan reducir los riesgos asociados. La formación ética es crucial para capacitar desarrolladores y otros profesionales del área. Cuando se incorpora entrenamiento en prácticas responsables, los especialistas logran entender mejor cómo el sesgo en los conjuntos de datos afecta considerablemente el funcionamiento de los sistemas inteligentes [31].

2.2.17 Perspectivas Tecnológicas y Sociales

2.2.18 Teoría de la Interacción Humano-Computadora (HCI)

La teoría HCI[32] se enfoca en desarrollar y evaluar sistemas computacionales que facilitan la interacción entre personas y máquinas. Entender cómo los LLMs funcionan eficientemente en diferentes aplicaciones es central para esta teoría, desde asistencia médica hasta educación e investigación científica.

La usabilidad forma parte de los principios básicos de HCI y se refiere a qué tan fácil resulta para los usuarios interactuar con un sistema específico. Con los LLMs, la usabilidad cambia dependiendo de cómo se organizan los estímulos y las respuestas que genera el modelo. Cuando el diseño de interfaz permite a los usuarios hacer preguntas claras y directas, la calidad de las respuestas mejora notablemente. Examinar cómo interactúan usuarios y LLMs, además de identificar qué elementos de diseño optimizan esta interacción, es fundamental para avanzar en el campo.

HCI reconoce la importancia de la retroalimentación en los procesos interactivos. Los LLMs necesitan ofrecer respuestas precisas, comprensibles y relevantes para cada usuario - capacidad fundamental en su funcionamiento. Los modelos requieren entrenamiento tanto en generación de texto como en adaptar sus respuestas al contexto específico y necesidades particulares de cada usuario. El feedback continuo de los usuarios opera como herramienta valiosa para mejorar la calidad de las interacciones, donde incrementar la eficacia de los LLMs mediante tales comentarios constituye una estrategia clave para optimizar su rendimiento en aplicaciones reales.

HCI destaca la relevancia de la ética al desarrollar sistemas interactivos. Las implicaciones éticas del uso de LLMs resultan cruciales conforme estos se integran en la vida diaria, abarcando privacidad de datos, transparencia en decisiones y reducción de sesgos inherentes. Quienes diseñan y desarrollan LLMs requieren mantenerse conscientes de tales desafíos mientras se enfocan en crear sistemas eficaces, equitativos y responsables[32]. .

2.2.19 Innovación Disruptiva

Clayton Christensen desarrolló su teoría de innovación disruptiva[33] para describir cómo nuevas tecnologías pueden revolucionar industrias completas y mercados establecidos. El proceso inicia cuando aparecen productos o servicios aparentemente inferiores a las opciones dominantes, aunque con el tiempo logran desplazarlas completamente. Los LLMs ejemplifican perfectamente esta dinámica, transformando las formas tradicionales de acceder y procesar información en campos tan diversos como educación, investigación científica y atención médica.

Los LLMs constituyen una innovación disruptiva que democratiza el acceso a vastos volúmenes de información mientras acelera considerablemente la creación de contenido. En contextos académicos, estos modelos apoyan a investigadores en tareas como sintetizar literatura especializada, reconocer patrones emergentes y desarrollar hipótesis nuevas, optimizando así los procesos de investigación y desarrollo. Las metodologías convencionales, por el contrario, operan con mayor lentitud y demandan recursos más extensos.

Los LLMs también amplían el acceso informativo para cualquier persona con disponibilidad tecnológica, permitiendo obtener respuestas coherentes y pertinentes expresadas en lenguaje natural. Esta capacidad empodera a individuos y comunidades previamente excluidos del conocimiento especializado[34].

2.2.20 Teoría de la Evaluación de Originalidad.

La Teoría de la Evaluación de Originalidad se centra en identificar y cuantificar la originalidad dentro de la investigación académica y científica. Cobra particular relevancia en el contexto de los LLMs, tecnologías que pueden utilizarse para validar la innovación presente en trabajos académicos y contribuir a los procesos de revisión por pares[34].

El criterio de originalidad constituye un elemento fundamental en la investigación, puesto que funciona como indicador de la calidad y el valor inherente a una contribución científica. En el ámbito educativo, validar la originalidad implica el examen del contenido de una obra mediante comparación con la literatura existente. Esta práctica puede incluir la identificación de

conceptos novedosos, metodologías innovadoras y la capacidad de un equipo investigador para aportar conocimientos inéditos al campo de estudio. Los LLMs, después de su entrenamiento con extensos corpus textuales, pueden utilizarse para ejecutar análisis comparativos entre una investigación y otros documentos relevantes. Asimismo, contribuyen a identificar similitudes y divergencias que pueden incidir en la valoración de la originalidad[34].

2.2.21 Modelos de Precios en Servicios en la Nube

Los modelos de precios de la nube se concentran en los métodos mediante los cuales las empresas establecen tarifas y estructuras de facturación para servicios en la nube. Con el crecimiento de las aplicaciones en la nube, estos modelos de precios experimentan mayor complejidad, evolucionando desde tarifas fijas hasta esquemas de facturación diversificados.

Esta perspectiva examina cómo los patrones de precios influyen en las decisiones del consumidor y la competitividad del mercado. Los resultados evidencian que la transparencia en la facturación y la confianza hacia los proveedores constituyen factores determinantes en la satisfacción del cliente.

De igual manera, se sugiere implementar un modelo de confianza que valide los eventos de facturación, estrategia que puede optimizar la percepción del cliente respecto a la imparcialidad y precisión de los procesos de cobro. Comprender estos patrones representa un aspecto fundamental para las organizaciones que buscan optimizar sus estrategias de precios y mejorar la experiencia del cliente en entornos basados en la nube[6].

2.2.22 Evaluación Empírica en Ingeniería de Software

Se enfoca en la necesidad de construir fundamentos científicos sólidos para la validación de técnicas basadas en LLM dentro de la Ingeniería de Software. Argumenta que la validación debe adherirse a prácticas de análisis científico claramente definidas, incluyendo el empleo de estadísticas inferenciales paramétricas y no paramétricas, elementos esenciales para obtener conclusiones científicas válidas ante la presencia de resultados no deterministas y con ruido[35].

Constituir un marco riguroso representa un aspecto fundamental para garantizar que las aplicaciones de LLM en la generación de pruebas y otras tareas de software sean validadas de manera exhaustiva y confiable.

2.2.23 Confianza en la Tecnología

Mientras la tecnología de LLMs como ChatGPT se vuelve más accesible, crece el interés en su aplicación legal. La investigación muestra que[36], aunque pocos participantes han usado LLMs para asesoría legal, un 45 % expresó interés en usarlos en el futuro, especialmente en áreas como tenencia, tráfico y derecho tributario. Usar estas herramientas como apoyo jurídico es una tendencia emergente que merece atención académica especializada.

Los participantes muestran cautela considerable hacia estos sistemas. Muchos reconocen que las respuestas de los LLMs requieren escepticismo, ya que pueden generar información imprecisa. Los usuarios necesitan desarrollar habilidades críticas para evaluar la calidad de lo que reciben. La confianza no depende solo de qué tan efectivos parecen estos modelos, también influye la experiencia previa y la habilidad para diferenciar respuestas confiables de las que no lo son. Desarrollar esta capacidad de discernimiento resulta fundamental para usar estas tecnologías de manera responsable.

Los LLMs no reemplazan la consultoría jurídica especializada. Funcionan como herramienta preliminar que ayuda a usuarios sin formación legal a comprender aspectos básicos antes de buscar orientación profesional. Aunque estos modelos muestran utilidad en contextos jurídicos, es importante abordar las preocupaciones sobre calidad informativa y capacitación del usuario. Esto permite un uso responsable y efectivo de esta tecnología emergente. La investigación busca mejorar la confiabilidad de los LLMs y entender mejor su impacto en la práctica legal[36].

CAPÍTULO III

MATERIALES Y MÉTODOS

3.1 Tipo de Investigación

Los LLMs revolucionaron el Procesamiento de Lenguaje Natural al analizar y comprender texto de forma similar a los humanos. Estos modelos descomponen documentos complejos y extraen información clave, lo que abre enormes posibilidades para automatizar el análisis de textos legales como términos y condiciones de software [37]. Esta funcionalidad resulta fundamental para abordar la complejidad inherente de documentos legales, generalmente extensos y redactados en lenguaje técnico que dificulta su comprensión para personas sin formación jurídica especializada.

Esta investigación se enfoca en desarrollar una solución tecnológica que emplea LLMs para analizar automáticamente términos y condiciones de software. Este enfoque permite evaluar objetivamente la efectividad del sistema propuesto mediante métricas cuantificables. También se establecen relaciones causales entre la implementación de LLMs y su capacidad para identificar correctamente cláusulas relevantes en documentos legales.

3.1.1 Metodología

Este estudio adopta un enfoque cuantitativo para crear y evaluar un sistema que analiza automáticamente términos y condiciones de software mediante Modelos de Lenguaje Extenso (LLMs). Elegimos este enfoque para obtener resultados objetivos y medibles que permitan evaluar con precisión el funcionamiento y efectividad del sistema desarrollado.

La metodología cuantitativa se adapta especialmente bien a esta investigación por varios motivos:

- Permite definir métricas precisas para evaluar el rendimiento del sistema a través de indicadores numéricos específicos como la precisión, el recall y el F1-score.
- Facilita la comparación objetiva entre distintas configuraciones y parámetros del modelo.
- Posibilita la aplicación de análisis estadísticos para validar los resultados obtenidos.
- Garantiza la reproducibilidad de la investigación mediante procedimientos sistemáticos y documentados.

El modelo hugging-quants/llama-3.2-3b-instruct ofrece una base robusta para implementar el sistema, logrando un equilibrio ideal entre la precisión del análisis y la eficiencia computacional. Para medir el rendimiento del sistema, se empleará un conjunto de documentos de términos y condiciones de software, organizados por tipo. Además, se compararán los resultados del análisis automatizado con un estándar de referencia que se ha establecido a través de un análisis realizado en GEval.

3.1.2 Varibales de Estudio

Considerando que son aspectos clave donde mediremos y analizaremos el sistema de análisis automático de términos y condiciones de software mediante LLMs, se han considerado las siguientes variables:

Variable Independiente:

- **Implementación de Modelos de Lenguaje Extenso (LLMs)** para el análisis de términos y condiciones, evaluada a través de diferentes configuraciones y parámetros del modelo seleccionado (llama-3.2-3b-instruct).

Variables Dependientes:

- **Exactitud del análisis:** Capacidad del sistema para identificar correctamente las cláusulas relevantes en los documentos.
- **Velocidad de procesamiento:** Tiempo requerido para analizar completamente un documento de términos y condiciones.

- **Comprensibilidad:** Facilidad con que los usuarios pueden entender los resúmenes generados por el sistema.

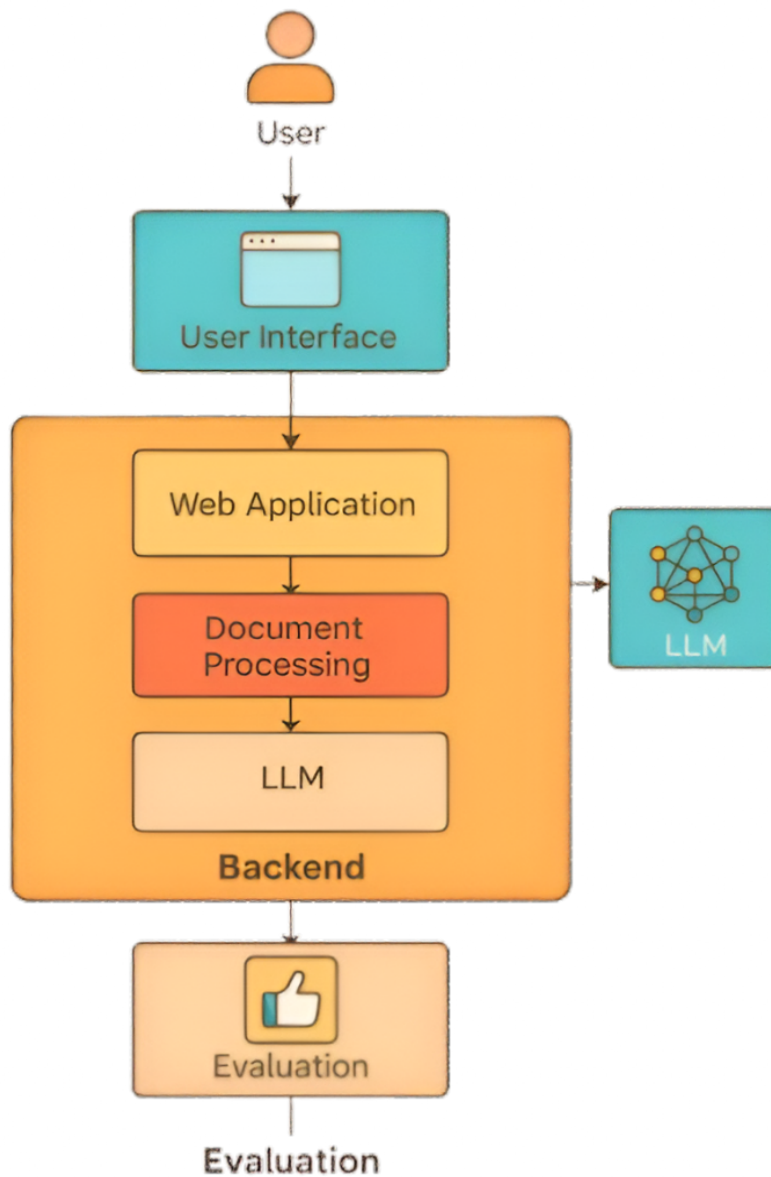


Fig. 3.
Diagrama de Procesos.

3.2 Población y Muestra

3.2.1 Población

La población de estudio comprende documentos de términos y condiciones de software comercial disponibles públicamente, tanto en español como en inglés. Estos documentos corresponden a diferentes categorías de productos y servicios tecnológicos, incluyendo servicios en

la nube, aplicaciones móviles, plataformas de colaboración, software de productividad, herramientas de desarrollo y otros servicios digitales.

3.2.2 Muestra

Para evaluar el sistema utilizamos una muestra de 19 documentos de términos y condiciones, seleccionados por conveniencia para asegurar diversidad en los tipos de software. La muestra se distribuye así:

- **Servicios en la nube:**

- AWS_Service_Terms
- AWS_Customer_Agreement
- MongoDB
- Mongo mx
- Postman

- **Plataformas de colaboración:**

- Zoom
- Microsoft Teams
- GitLab Privacy Statement | GitLab

- **Software de productividad:**

- Microsoft Office
- Adobe

- **Herramientas de desarrollo:**

- Visual Studio Code
- LM Studio Terms of Use

- **Servicios de diseño y educación:**

- Terms of Use Canva
- Udemey

- **Servicios multimedia y otros:**
 - Apple Media Services - Apple
 - Google Workspace
 - Hugging Face
 - MATLAB & Simulink
 - Meta Platforms Technologies complementarias

Para seleccionar los documentos aplicamos estos criterios:

1. Documento completo de acceso público
2. Relevancia para usuarios de software en entornos profesionales y educativos
3. Variedad de sectores tecnológicos incluidos
4. Documentos en español e inglés

Los 19 documentos de esta muestra son suficientes para evaluar la efectividad del sistema de análisis automatizado en diferentes contextos y tipos de documentos legales de software.

3.3 Diseño de la Investigación

3.3.1 Materiales, Equipos y Software

El diseño de investigación detalla los materiales, equipos y software necesarios para desarrollar efectivamente el proyecto. Estos recursos garantizan la precisión, relevancia y eficiencia del sistema automatizado de análisis de términos y condiciones de software con LLMs. Los elementos que conforman este diseño se describen a continuación.

La **Tabla 1** muestra de manera integral todos los materiales y recursos empleados en la implementación del sistema.

Tabla I.
RECURSOS Y HERRAMIENTAS UTILIZADOS

Categoría	Descripción	Justificación
Hardware	Servidor de desarrollo con mínimo 8GB RAM y procesador 4 núcleos	Necesario para entrenamiento y ejecución de modelos LLM
	Unidad de Procesamiento Gráfico (GPU) - NVIDIA con mínimo 4GB VRAM	Acelera significativamente el procesamiento de documentos y la inferencia del modelo
	Almacenamiento SSD de 512GB	Facilita el acceso rápido a documentos y modelos
Software	Python 3.9+	Lenguaje principal de desarrollo
	FastAPI	Framework para desarrollo de API RESTful
	LM Studio	Plataforma para gestión local de LLMs
	llama.cpp	Biblioteca optimizada para inferencia de modelos LLM
	Hugging Face Transformers	Biblioteca para acceso a modelos preentrenados
	PyMuPDF, python-docx, nltk, spaCy, BeautifulSoup, docx2txt, markdown	Procesamiento y análisis de documentos en diversos formatos
	React	Biblioteca para desarrollo de interfaces de usuario
Modelos LLM	Llama 3.2 3B Instruct	Modelo base seleccionado para el sistema
	hugging-quants/llama-3.2-3binstruct	Modelo optimizado para instrucciones específicas
Conjuntos de datos	Documentos de términos y condiciones de software	Documentos para evaluación y testing
	Preguntas frecuentes sobre términos legales	Para validación de respuestas
Herramientas de evaluación	Model-Based Scorers	Evaluación de relevancia y precisión

3.3.2 Entorno de desarrollo del sistema

Desarrollamos el sistema usando Python como lenguaje de programación principal por su versatilidad, facilidad de uso y amplia adopción en Procesamiento de Lenguaje Natural (NLP) e Inteligencia Artificial. Python ofrece múltiples bibliotecas especializadas para analizar documentos legales: spaCy y NLTK procesan texto, PyMuPDF y python-docx extraen información de documentos, mientras que scikit-learn clasifica y analiza patrones en textos legales [38].

Python también se integra fácilmente con LLMs mediante herramientas como Hugging Face Transformers, LangChain, llama.cpp y LM Studio, aplicando técnicas avanzadas de IA para

interpretar cláusulas de términos y condiciones. Su compatibilidad con APIs de modelos pre-entrenados y frameworks como FastAPI o Flask permite además desplegar eficientemente el sistema[39].

El proyecto requiere análisis legal, donde Python ofrece herramientas para detectar riesgos en términos y condiciones, analizar cumplimiento normativo (GDPR, CCPA) y asegurar la explicabilidad de modelos de IA. Estas capacidades convierten a Python en la opción ideal para este desarrollo.[40]

- **Ecosistema de IA maduro:** Python domina el campo de inteligencia artificial y procesamiento de lenguaje natural, con un ecosistema robusto de bibliotecas especializadas y frameworks optimizados para LLMs.
- **Facilidad de integración:** La versatilidad de Python facilita su integración con componentes tecnológicos diversos - sistemas de procesamiento de documentos, interfaces de usuario y otros elementos clave. Proyectos que manejan múltiples formatos de entrada y salida encuentran en esta característica un aspecto fundamental.
- **Comunidad activa:** Python cuenta con una amplia comunidad de desarrolladores que proporciona soporte continuo, documentación exhaustiva y actualizaciones frecuentes de bibliotecas. Esta característica permite la sostenibilidad a largo plazo del proyecto.

FastAPI funciona como framework moderno de alto rendimiento para crear APIs RESTful. Este framework aporta ventajas significativas al proyecto:

- **Alto rendimiento:** FastAPI supera en rendimiento a frameworks tradicionales como Flask o Django. Su diseño asíncrono nativo permite manejar múltiples solicitudes de procesamiento de documentos simultáneamente, aspecto fundamental para el proyecto.
- **Documentación automática:** FastAPI genera documentación interactiva automática (OpenAPI/Swagger y ReDoc). Esta documentación mejora la comprensión y prueba de endpoints, mientras acelera el desarrollo y simplifica futuras integraciones.
- **Validación de datos:** Pydantic permite la validación y serialización de datos en FastAPI. Esta implementación reduce significativamente el código necesario para validar entradas y asegura la integridad de los datos procesados.

- **Soporte para operaciones asíncronas:** FastAPI maneja operaciones extensas como el procesamiento de documentos sin bloquear el servidor. Esta capacidad mejora la eficiencia de recursos del sistema.

3.4 Metodología de desarrollo

Elegimos la metodología de desarrollo en cascada para crear el sistema automatizado de análisis de términos y condiciones, este enfoque estructurado y secuencial coincide con los objetivos académicos de la investigación, además facilita documentar cada etapa del proceso de forma clara. La metodología en cascada funciona de manera secuencial cada fase se completa antes de iniciar la siguiente. Cuenta con entregables definidos que alimentan las etapas posteriores. La Figura 4 muestra las cinco fases principales de la metodología adoptada:



Fig. 4.

Modelo de Cascada Adaptado para el Desarrollo del Sistema de Análisis Automatizado de Términos y Condiciones mediante LLMs.

3.4.1 Cronograma de Desarrollo

Siguiendo la metodología en cascada adoptada para este proyecto, se estableció un cronograma detallado que organiza las actividades en fases secuenciales. Este cronograma permite

una planificación estructurada del desarrollo del sistema, asegurando que cada fase se complete antes de iniciar la siguiente, conforme a los principios del modelo en cascada.

Fase 1: Análisis y definición de requisitos

Tabla II.
CRONOGRAMA DE ACTIVIDADES - FASE 1

Semana	Actividad	Descripción	H
Sem. 1	Revisión bibliográfica	Estudio de artículos sobre LLMs y NLP	7
	Análisis de tecnologías	Investigación de modelos open-source	6
	Exploración de LM Studio	Pruebas iniciales con la plataforma	7
	Definición del problema	Documentación del problema a resolver	5
Sem. 2	Requisitos funcionales	Documentación de funcionalidades	10
	Requisitos no funcionales	Especificación de rendimiento y seguridad	7
	Documento de especificaciones	Compilación de requisitos	5

Fase 2: Diseño del sistema

Tabla III.
CRONOGRAMA DE ACTIVIDADES - FASE 2

Semana	Actividad	Descripción	H
Sem. 3	Arquitectura general	Diseño de arquitectura de 3 capas	5
	Diseño de componentes	Definición de módulos y sus interacciones	6
	Flujo de datos	Diagramas de flujo de información	6
	Selección de tecnologías	Justificación de tecnologías utilizadas	6
Sem. 4	Estructura de directorios	Organización del código	5
	Diseño de interfaces	Wireframes y mockups de UI	4
	Diseño de API	Especificación de endpoints y formatos	6
	Diseño de caché	Sistema de almacenamiento temporal	7

Fase 3: implementación

Tabla IV.
CRONOGRAMA DE ACTIVIDADES - FASE 3

Semana	Actividad	Descripción	H
Sem. 5	Entorno de desarrollo	Instalación de Python, Node.js, dependencias	8
	Estructura base del proyecto	Creación de directorios y archivos iniciales	7
	Configuración de FastAPI	Setup inicial del framework y rutas básicas	8
	Configuración de LM Studio	Instalación y pruebas del modelo Llama 3.2 3B	7
Sem. 6	Procesador de documentos	Desarrollo de funciones para PDF, DOCX, TXT	6
	Sistema de caché	Implementación de almacenamiento temporal	7
Sem. 7	LLM Service	Integración con LM Studio y prompts optimizados	7
	Sistema de análisis	Algoritmos para identificación de cláusulas	7
Sem. 8	Frontend base	Estructura HTML/CSS con Bootstrap	5
	Sistema de chat	Implementación de interfaz conversacional	6
	Integración frontend-backend	Conexión de servicios y manejo de respuestas	6

Fase 4: Verificación y pruebas

Tabla V.
CRONOGRAMA DE ACTIVIDADES - FASE 4

Semana	Actividad	Descripción	H
Sem. 9	Pruebas de procesador	Verificación de extracción de texto en todos los formatos	6
	Pruebas de LLM service	Validación de respuestas y tiempos de respuesta	5
	Pruebas de API	Testing de todos los endpoints	4
Sem. 10	Corpus de evaluación	Preparación de 19 documentos de prueba	6
	Implementación GEval	Script de evaluación automatizada	6
	Análisis de resultados	Métricas de precisión, recall, fidelidad	5
	Corrección de errores	Ajustes basados en resultados de pruebas	5

Fase 5: Implementación y mantenimiento

Tabla VI.
CRONOGRAMA DE ACTIVIDADES - FASE 5

Semana	Actividad	Descripción	H
Sem. 11	Configuración de servidor	Setup en macOS Sequoia con 8GB RAM	8
	Optimización de rendimiento	Ajustes de concurrencia y caché	7
	Pruebas de producción	Verificación en entorno final	6
	Monitoreo inicial	Configuración de GEval para métricas	5
Sem. 12	Manual de usuario	Guía completa con capturas de pantalla	5
	Manual técnico	Documentación de instalación y mantenimiento	5
	Informe final	Compilación de resultados y conclusiones	3
	Presentación	Preparación y ensayo de defensa del proyecto	7

3.4.2 Fase de Análisis y Definición de Requisitos

En esta fase inicial se realizó un análisis de los diferentes LLMs, con énfasis en aquellos que son de código abierto, así mismo se utilizó LM Studio ya que facilita la ejecución de LLMs localmente.

Investigación Bibliográfica Se revisaron artículos científicos, documentación técnica y estudios comparativos sobre modelos LLM, centrándose en su capacidad para procesar y comprender textos legales de lo cual se puede ver en la sección 2.2 Bases teóricas de este documento.

Definición de criterios de selección:

- Capacidad para procesar documentos en múltiples formatos (PDF, DOCX, TXT)
- Precisión en la comprensión de lenguaje jurídico
- Extracción de derechos y obligaciones del usuario
- Generación de resúmenes comprensibles en lenguaje sencillo
- Eficiencia computacional (requisitos de hardware)
- Capacidad multilingüe (español e inglés)

Definición de requisitos no funcionales:

- Procesamiento eficiente (menos de 2 minutos por documento)
- Precisión superior al 85 % en la identificación de cláusulas
- Capacidad para manejar documentos extensos (hasta 100 páginas)
- Compatibilidad con hardware moderado (8GB RAM)
- Soporte para documentos en español e inglés

Documentación de requisitos:

- Se elaboró un documento detallado de especificación de requisitos como entregable principal de esta fase, que sirvió como base para las siguientes etapas del desarrollo.

3.4.3 Fase de diseño del sistema

Después de completar el análisis y definición de requisitos, desarrollamos el diseño detallado de la arquitectura del sistema. Establecer un marco sólido resultaba fundamental para implementar eficazmente el sistema de análisis automático de términos y condiciones con LLMs.

Diseño de la arquitectura del sistema

Diseñamos una arquitectura modular de múltiples capas basándose en principios de separación de responsabilidades y alta cohesión/bajo acoplamiento. Esta arquitectura incluye las siguientes capas:

Capa de presentación (Frontend)

- **Interfaz web:** HTML5, CSS3 y JavaScript componen la base de desarrollo. Bootstrap proporciona el diseño responsivo.
- **Sistema de plantillas:** Jinja2 genera contenido HTML dinámico en el sistema.
- **Interacción del usuario:** La interfaz conversacional simula un chat para interactuar naturalmente con el sistema.
- **Componentes principales:**
 - Sistema de chat interactivo con indicadores de escritura.
 - Visualización de análisis y resúmenes generados por el LLM.
 - Notificaciones tipo toast para informar al usuario sobre el estado de las operaciones.
 - Pantallas modales para procesos en segundo plano.

Capa de API (Backend)

- **Framework:** FastAPI fue seleccionado por su rendimiento, documentación automática y soporte para operaciones asíncronas.
- **Estructura de endpoints:**
 - Endpoints RESTful para todas las operaciones.
 - Validación automática de parámetros mediante modelos Pydantic.

- Gestión de errores centralizada con códigos de estado HTTP apropiados.
- Soporte para carga de archivos con validación de tamaño y formato.

Capa de servicios:

- **document_processor** Encargado de la extracción y procesamiento de texto de diferentes formatos de archivo.
- **llm_service** Responsable de la interacción con el modelo Llama 3.2 3B Instruct y el procesamiento inteligente del texto.
- Cada servicio encapsula una funcionalidad específica, facilitando el mantenimiento y las pruebas.

Capa de almacenamiento:

Se diseñó un sistema de almacenamiento basado en archivos para:

- Documentos cargados (carpeta uploads/).
- Caché de resultados (carpeta 'cache/')

3.4.4 Diseño detallado de la estructura de directorios



Fig. 5.
Estructura de Directorios: Análisis Automático con LLMs

Esta estructura se diseñó siguiendo el principio de “convención sobre configuración”, donde cada componente tiene una ubicación predecible según su función, facilitando la navegación y mantenimiento del código.

3.4.5 Requisitos Funcionales

Siguiendo el formato de la referencia, presento los requisitos funcionales del sistema:

Tabla VII.
RF001 CARGA DE DOCUMENTOS

Código:	Nombre	Prioridad
RF001	Carga de documentos de T&C	Alta
Descripción	Funcionalidad para cargar documentos de términos y condiciones en múltiples formatos	
Proceso	<ul style="list-style-type: none"> - El usuario podrá cargar documentos en formatos PDF, DOCX y TXT - El sistema validará el formato y tamaño del archivo (máximo 50MB) - Se mostrará una barra de progreso durante la carga - Se generará un ID único para cada documento cargado - El sistema almacenará el documento en el directorio /uploads/ - Se notificará al usuario cuando la carga sea exitosa o fallida 	

Tabla VIII.
RF002 ANÁLISIS AUTOMÁTICO

Código:	Nombre	Prioridad
RF002	Análisis automático de documentos	Alta
Descripción	Procesamiento automático de documentos para identificar cláusulas clave	
Proceso	<ul style="list-style-type: none"> - El sistema extraerá el texto del documento cargado - Se aplicará preprocesamiento del texto (normalización, eliminación de caracteres especiales) - El modelo LLM identificará cláusulas de: privacidad, pagos, cancelación, responsabilidad - Se generará un resumen en lenguaje sencillo - Los resultados se almacenarán en caché para consultas futuras - El tiempo máximo de procesamiento será de 2 minutos 	

Tabla IX.
RF003 INTERFAZ DE CHAT

Código:	Nombre	Prioridad
RF003	Interfaz conversacional	Alta
Descripción	Sistema de chat para realizar consultas sobre el documento	
Proceso	<ul style="list-style-type: none"> - El usuario podrá escribir preguntas en lenguaje natural - El sistema procesará la pregunta usando el modelo LLM - Se generarán respuestas contextualizadas basadas en el documento - El tiempo de respuesta será inferior a 30 segundos - Se mantendrá un histórico de la conversación - Se mostrarán indicadores visuales mientras se procesa la respuesta 	

Tabla X.
RF004 VISUALIZACIÓN DE RESULTADOS

Código:	Nombre	Prioridad
RF004	Visualización de análisis	Alta
Descripción	Presentación estructurada de los resultados del análisis	
Proceso	<ul style="list-style-type: none"> - Se mostrarán las cláusulas categorizadas por tipo - Se destacarán visualmente términos críticos - Se generará una versión simplificada del documento - El usuario podrá alternar entre vista detallada y simplificada - Se permitirá exportar los resultados en formato PDF - Se utilizarán colores para indicar el nivel de importancia de las cláusulas 	

Tabla XI.
RF005 GESTIÓN DE SESIONES

Código:	Nombre	Prioridad
RF005	Gestión de sesiones de usuario	Media
Descripción	Control de sesiones para mantener el contexto del análisis	
Proceso	<ul style="list-style-type: none"> - Se generará un ID de sesión único para cada documento cargado - La sesión mantendrá el estado del análisis y el historial de chat - El usuario podrá eliminar la sesión cuando termine - Los archivos temporales se limpiarán automáticamente - Se implementará un timeout de sesión por inactividad 	

3.4.6 Requisitos No Funcionales

Tabla XII.
RNF-USA-002 DISEÑO RESPONSIVO

Código:	Nombre	Prioridad
RNF-USA-002	Diseño responsivo	Alta
Descripción	Adaptación a diferentes dispositivos	
Proceso	La aplicación web utilizará Bootstrap para garantizar una visualización adecuada en computadores, tablets y teléfonos móviles	

Tabla XIII.
RNF-REN-001 RENDIMIENTO

Código:	Nombre	Prioridad
RNF-REN-001	Rendimiento del sistema	Alta
Descripción	Tiempos de respuesta y procesamiento	
Proceso	<ul style="list-style-type: none">- Tiempo de carga de página <3 segundos- Análisis de documento <45 segundos- Respuesta de chat <30 segundos- Soporte para 5 usuarios concurrentes	

Tabla XIV.
RNF-SEG-001 SEGURIDAD DE DATOS

Código:	Nombre	Prioridad
RNF-SEG-001	Seguridad y privacidad	Alta
Descripción	Protección de datos y documentos	
Proceso	<ul style="list-style-type: none">- Los documentos se procesarán localmente- No se almacenarán datos personales- Los archivos se eliminarán después de la sesión- Comunicación segura mediante HTTPS	

3.4.7 Fase de implementación

Implementación del procesador de documentos

Se desarrolló el módulo `document_processor.py`, un sistema robusto para el procesamiento de documentos en diferentes formatos, incorporando las siguientes funcionalidades:

- **Extracción de texto multiformato:** Se implementaron métodos específicos para la extracción eficiente de texto de archivos PDF, DOCX, TXT, utilizando las bibliotecas PyMuPDF, python-docx y BeautifulSoup respectivamente.
- **Preprocesamiento de texto:** Se desarrollaron algoritmos para mejorar la calidad del análisis, incluyendo:
 - Normalización de espacios y caracteres especiales
 - Eliminación de cabeceras y pies de página redundantes
 - Corrección de errores comunes en la extracción de PDFs
 - Preservación de la estructura jerárquica del documento
- **Segmentación inteligente:** Se implementó un sistema de chunking que divide el documento en segmentos manejables para el procesamiento con LLMs, considerando:
 - Preservación de la integridad semántica de secciones y párrafos
 - Manejo de referencias cruzadas entre secciones
 - Adaptación dinámica del tamaño de los chunks según la complejidad del texto
- **Manejo de documentos extensos:** Para documentos de gran tamaño, se implementó un procesamiento paralelo que optimiza la utilización de recursos del sistema, reduciendo significativamente los tiempos de análisis.

El siguiente fragmento de código ilustra la implementación central del procesador de documentos:

```

class DocumentProcessor:
    def __init__(self, config=None):
        self.config = config or default_config
        self.supported_formats = {
            '.pdf': self._process_pdf,
            '.docx': self._process_docx,
            '.txt': self._process_txt,
        }

    async def process_document(self, file_path):
        """Procesa un documento y extrae su contenido como texto."""
        ext = os.path.splitext(file_path)[1].lower()

        if ext not in self.supported_formats:
            raise UnsupportedFormatError(f"Formato no soportado: {ext}")

        text = await self.supported_formats[ext](file_path)
        preprocessed_text = self._preprocess_text(text)
        chunks = self._create_chunks(preprocessed_text)

        return {
            'full_text': preprocessed_text,
            'chunks': chunks,
            'metadata': self._extract_metadata(file_path, text)
        }

```

Fig. 6.

Procesador de documentos.

Integración con LM Studio

Se implementó la conexión con Llama 3.2 3B Instruct a través de LM Studio, configurando los parámetros óptimos para el análisis de textos legales:

- **Inicialización y configuración:** Se estableció una comunicación con la API local de LM Studio, configurando parámetros fundamentales:
- **Inicialización y configuración:** Se estableció una comunicación con la API local de LM Studio, configurando parámetros fundamentales:
 - Temperatura: 0.7 para equilibrar creatividad y precisión
 - Top-p: 0.9 para mantener respuestas coherentes
 - Tamaño máximo de tokens: 4096 para permitir análisis de secciones extensas
 - Contexto máximo: 16384 para mantener coherencia en el análisis

```

# Configuración de LMStudio
LMSTUDIO_API_URL: str = os.getenv("LMSTUDIO_API_URL", "http://localhost:1234/v1")

# Modelo (cuantizado para mejor rendimiento)
LMSTUDIO_MODEL: str = os.getenv("LMSTUDIO_MODEL", "hugging-quants/llama-3.2-3b-instruct")

# Parámetros del modelo – optimizados para velocidad
MODEL_TEMPERATURE: float = 0.65 # Más bajo para respuestas más deterministas
MODEL_MAX_TOKENS: int = 800 # Reducido para respuestas más rápidas
MODEL_MAX_CONTEXT_SIZE: int = 8192 # Tamaño máximo de contexto para el modelo
MODEL_FREQUENCY_PENALTY: float = 0.1 # Reducido para mayor velocidad
MODEL_PRESENCE_PENALTY: float = 0.1 # Reducido para mayor velocidad

```

Fig. 7.

Implementación de LM Studio.

- **Gestión de prompts:** Se desarrolló un sistema de plantillas de prompts diseñadas específicamente para el análisis de documentos legales, que incluyen:
 - Instrucciones para identificación de cláusulas clave
 - Pautas para extraer derechos y obligaciones
 - Directrices para generar resúmenes comprensibles
- **Optimización de recursos:** Se implementaron estrategias para maximizar la eficiencia:
 - Sistema de caché para evitar procesamiento
 - Control de concurrencia para optimizar la utilización de la GPU
 - Mecanismos de timeout para gestionar consultas complejas

Desarrollo del sistema de análisis

Los componentes de análisis automático de términos y condiciones extraen y estructuran información relevante:

- **Identificación de cláusulas:** Implementación de algoritmos para detectar y clasificar automáticamente distintos tipos de cláusulas, incluyendo:
- **Inicialización y configuración:** Se estableció una comunicación con la API local de LM Studio, configurando parámetros fundamentales:
 - Derechos del usuario
 - Obligaciones y restricciones

- Limitaciones de responsabilidad
- Políticas de privacidad
- Términos de cancelación y reembolso
- **Análisis redundante:** Desarrollo de técnicas para comprender el significado contextual de las cláusulas y evaluar su impacto potencial para el usuario.
- **Generación de resúmenes:** Implementación de métodos para crear resúmenes claros y concisos de los términos y condiciones, adaptados a diferentes niveles de comprensión.

Implementación de la interfaz de usuario

Se desarrolló una interfaz web utilizando HTML, CSS y JavaScript, con un enfoque en la usabilidad y la experiencia del usuario:

- **Diseño responsivo:** La interfaz se adapta a diferentes dispositivos y tamaños de pantalla, garantizando una experiencia consistente.
- **Carga de documentos:** Implementación de un sistema intuitivo para la carga de documentos con validación en tiempo real de formatos y tamaños.
- **Visualización de resultados:** Desarrollo de componentes para mostrar de manera clara y organizada:
 - Resúmenes generales del documento
 - Categorización de cláusulas relevantes
 - Derechos y obligaciones destacados
 - Términos potencialmente problemáticos
- **Interacción conversacional:** Implementación de una interfaz de chat que permite a los usuarios formular preguntas en lenguaje natural sobre el documento.

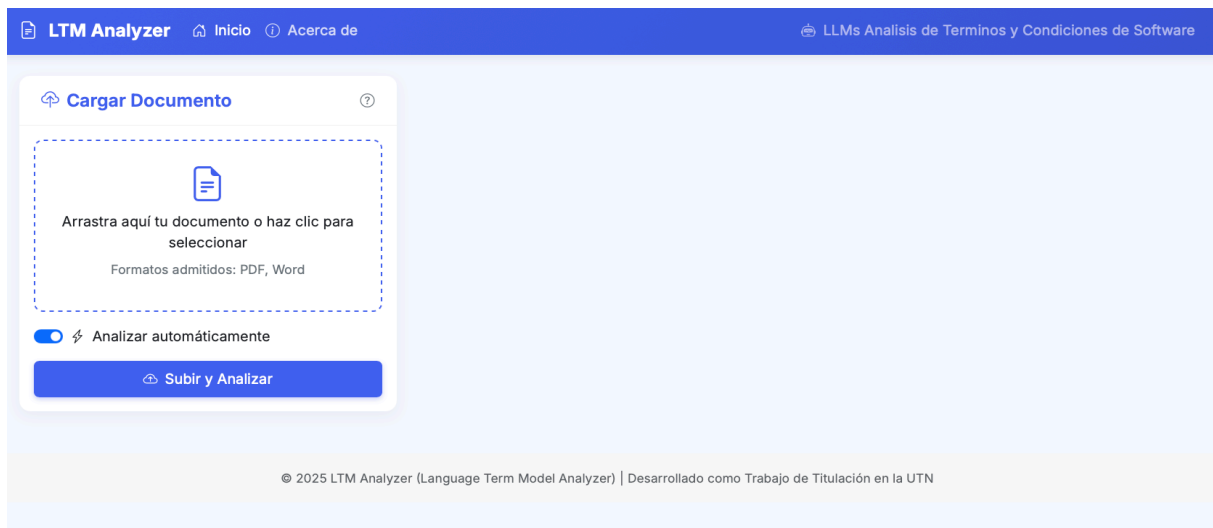


Fig. 8.
Interfaz de usuario del sistema.

3.4.8 Fase de verificación y pruebas

Al subir el documento sea pdf, word o txt al modelo al realizarse preguntas sobre términos y condiciones nos da como resultado un archivo .json con la pregunta realizada y con la respuesta del modelo.

```
{
  "timestamp": 1743124051.08753,
  "query": "¿Cuáles son las obligaciones del proveedor hacia el usuario según estos términos?",
  "result": {
    "success": true,
    "content": "Estas son las obligaciones del proveedor (Google) hacia el usuario:\n\n* Proporcionar los Servicios de conformidad con el Acuerdo.\n* Mantener la seguridad de las cuentas y datos del usuario.\n* Notificar al usuario al menos 12 meses antes de discontinuar cualquier Servicio Principal.\n* Realizar actualizaciones comercialmente razonables de los Servicios.\n* Cumplir con la ley aplicable, abordar riesgos importantes de seguridad o evitar cargas económicas o técnicas sustanciales.\n* Facturar al usuario de manera justa y transparente (con opciones de facturación flexibles).\n* Proporcionar soporte técnico adecuado.\n\n¿Esto responde a tu pregunta?",
    "usage": {
      "prompt_tokens": 2162,
      "completion_tokens": 148,
      "total_tokens": 2310
    }
  },
  "elapsed_time": 15.3495380878448
}
```

Fig. 9.
Ejemplo de salida JSON con respuesta del modelo a una consulta específica sobre términos y condiciones.

Una vez completada la implementación del sistema, se procedió a realizar una verificación exhaustiva de su funcionamiento y a evaluar su rendimiento en el análisis de términos y condiciones de software,

Planificación de pruebas

Se desarrolló un plan detallado de pruebas que abarcó diferentes aspectos del sistema:

- **Pruebas unitarias:** Para verificar el funcionamiento correcto de cada componente individual.
- **Pruebas de integración:** Para comprobar la interacción adecuada entre los distintos módulos.
- **Pruebas de sistema:** Para evaluar el funcionamiento integral de la aplicación.
- **Pruebas de rendimiento:** Para medir tiempos de respuesta y uso de recursos.

Comparación de modelos LLM evaluados

Elegir el modelo LLM más adecuado para analizar términos y condiciones de software representaba un paso crítico del desarrollo. Evaluamos varios modelos de código abierto disponibles localmente a través de LM Studio para asegurar una elección informada que equilibrara rendimiento y eficiencia. Esta evaluación siguió principios teóricos de la literatura científica actual [13].

Fundamentos teóricos para la evaluación

Zhao [13] clasifican los LLMs en tres categorías según su tamaño: pequeños (1-10B parámetros), medianos (10-100B parámetros) y grandes (>100B parámetros). Cada categoría presenta características distintivas que influyen en su aplicabilidad a tareas específicas, como se muestra en la siguiente tabla XVI:

Tabla XV.
CLASIFICACIÓN DE MODELOS DE LENGUAJE POR CATEGORÍA DE TAMAÑO[13]

Categoría de tamaño	Parámetros	Ejemplos de modelos	Capacidades principales	Limitaciones
Pequeños	1–10B	LLaMA-7B, Phi-1, Mistral-7B	Tareas básicas de NLP, instrucciones simples, generación de texto	Razonamiento limitado, conocimiento especializado reducido
Medianos	10–100B	LLaMA65B, Falcon40B, MPT30B	Razonamiento multi-etapa, seguimiento de instrucciones complejas	Mayor costo computacional, latencia moderada
Grandes	>100B	GPT-4, PaLM, Claude	Capacidades emergentes, razonamiento complejo, amplio conocimiento especializado	Alto costo computacional, dificultad de despliegue, latencia alta

Además, Zhao [13]] destacan que existen múltiples arquitecturas para estos modelos, cada una con características distintivas que afectan su rendimiento en diferentes tipos de tareas:

Tabla XVI.
ARQUITECTURAS POPULARES DE LLMS [13]

Arquitectura	Descripción	Ventajas	Desventajas	Ejemplos
Decoder-only	Generación de texto autoregresiva	Eficiente para generación de texto	Contexto unidireccional	GPT, LLaMA, PaLM
Encoder-decoder	Codificación bidireccional con decodificación autoregresiva	Mejor comprensión y generación	Mayor complejidad computacional	T5, BART, Flan-T5
Encoder-only	Representación bidireccional del contexto	Excelente comprensión de texto	Limitada capacidad generativa	BERT, RoBERTa, DeBERTa

El survey también señala que los modelos con fine-tuning en instrucciones muestran mejoras significativas respecto a los modelos base, como se detalla en la siguiente tabla VIII:

Tabla XVII.
IMPACTO DEL FINE-TUNING EN INSTRUCCIONES[13]

Aspecto	Modelo base	Modelo con fine-tuning en instrucciones
Seguimiento de instrucciones	Limitado	Significativamente mejorado
Precisión en tareas específicas	Moderada	Alta
Adaptabilidad a formatos de salida	Baja	Alta
Reducción de respuestas irrelevantes	Baja	Alta
Capacidad conversacional	Limitada	Mejorada

Metodología de evaluación

Los benchmarks y métodos de evaluación propuestos [13] guiaron el diseño de un protocolo de evaluación. Este protocolo considera las capacidades clave para el análisis de documentos legales:

Los benchmarks mencionados por Zhao [13] no se diseñaron específicamente para evaluación de análisis legal, pero sus principios ayudaron a definir criterios de evaluación práctica. Desarrollamos un conjunto de pruebas específicas con 5 documentos representativos de términos y condiciones, evaluando:

Modelos evaluados

Tabla XVIII.
EVALUACIÓN DE CAPACIDADES ESPECÍFICAS DE LLMS[13]

Capacidad	Descripción	Benchmarks relevantes	Importancia para análisis legal
Comprensión de texto	Capacidad para entender el significado del texto	GLUE, SuperGLUE, MMLU	Alta (interpretación de cláusulas)
Razonamiento	Habilidad para realizar inferencias lógicas	GSM8K, MATH, BBH	Alta (análisis de implicaciones)
Conocimiento factual	Precisión en información recuperada	TriviaQA, Natural-Questions	Media (términos técnicos)
Seguimiento de instrucciones	Capacidad para seguir indicaciones precisas	HELM, AlpacaEval	Alta (identificación de cláusulas)
Veracidad	Tendencia a proporcionar información factual correcta	TruthfulQA, FACTOR	Crítica (precisión legal)

Aplicando el marco teórico de clasificación de modelos propuesto por Zhao [13], que establece categorías basadas en tamaño (pequeños: 1-10B, medianos: 10-100B, grandes: >100B) y arquitectura (decoder-only, encoder-decoder, encoder-only), seleccionamos para nuestra investigación cinco modelos actuales que representan diferentes combinaciones de estas características. La selección de estos modelos específicos constituye una contribución original de este estudio, considerando modelos de reciente desarrollo que se ajustan a los criterios teóricos establecidos en el survey:

Tabla XIX.
MODELOS SELECCIONADOS PARA LA EVALUACIÓN

Modelo	Tamaño	Desarrollador	Características	Categoría según Zhao et al. [14]
Llama 3.2 3B Instruct	3B	Meta	Optimizado para instrucciones, bajo requerimiento de recursos	Pequeño, decoder-only, instruction-tuned
Llama 3 8B	8B	Meta	Mayor capacidad, requiere más recursos	Pequeño, decoder-only
Mistral 7B	7B	Mistral AI	Buen rendimiento en tareas de comprensión	Pequeño, decoder-only
Phi-3 Mini	3.8B	Microsoft	Eficiente en recursos, buen razonamiento	Pequeño, decoder-only
OpenChat 3.5	7B	OpenChat Team	Especializado en diálogos conversacionales	Pequeño, decoder-only, fine-tuned

Estos modelos fueron seleccionados específicamente para nuestra investigación por estar disponibles en LM Studio y representar diferentes puntos en el espectro de tamaño, eficiencia y capacidades, permitiendo una comparación empírica que valide los principios teóricos descritos por Zhao [13] en el contexto específico del análisis de términos y condiciones de software.

Resultados comparativos

La evaluación de esta investigación reveló diferencias notables en el rendimiento de los modelos seleccionados. Los hallazgos contribuyen al campo porque no existen evaluaciones previas de estos modelos específicos para analizar términos y condiciones. Los resultados provienen de diversas métricas diseñadas para medir la relevancia en el análisis de documentos legales lo que representa en las siguientes Tablas.

Tabla XX.
PRECISIÓN EN IDENTIFICACIÓN DE CLÁUSULAS CLAVE (0-100 %)

Modelo	Precisión	Observaciones
Llama 3.2 3B Instruct	84.5 %	Excelente capacidad para identificar derechos y obligaciones contractuales
Llama 3 8B	86.3 %	Ligera mejora respecto al modelo de 3B, pero con mayor costo computacional
Mistral 7B	79.2 %	Buen rendimiento general, algunas dificultades con terminología técnica especializada
Phi-3 Mini	76.8 %	Rendimiento aceptable considerando su tamaño compacto
OpenChat 3.5	81.4 %	Fortaleza en procesamiento dialógico, precisión reducida en documentos extensos

Tabla XXI.
TIEMPO DE PROCESAMIENTO PROMEDIO (SEGUNDOS POR DOCUMENTO)

Modelo	Tiempo (s)	Observaciones
Llama 3.2 3B Instruct	42.3	Excelente equilibrio entre velocidad y precisión
Llama 3 8B	78.6	Mayor latencia debido al incremento de parámetros
Mistral 7B	65.8	Rendimiento intermedio
Phi-3 Mini	38.5	Procesamiento más rápido, con compromiso en precisión
OpenChat 3.5	71.2	Procesamiento relativamente lento

Requisitos de memoria (RAM mínima recomendada)

Tabla XXII.
REQUISITOS COMPUTACIONALES POR MODELO

Modelo	RAM (GB)	Observaciones
Llama 3.2 3B Instruct	8 GB	Operación eficiente en hardware de especificaciones moderadas
Llama 3 8B	16 GB	Requiere infraestructura de gama media-alta
Mistral 7B	12 GB	Requisitos intermedios
Phi-3 Mini	8 GB	Optimizado para hardware limitado
OpenChat 3.5	16 GB	Alto consumo de recursos computacionales

Estos resultados empíricos pueden interpretarse mejor considerando las técnicas de eficiencia para modelos de lenguaje descritas por Zhao [13]. Varios de los modelos evaluados en nuestra investigación incorporan estas técnicas, lo que explica sus características distintivas de rendimiento:

Tabla XXIII.
TÉCNICAS DE EFICIENCIA PARA MODELOS DE LENGUAJE SEGÚN ZHAO [13]

Técnica	Descripción	Beneficios	Impacto en rendimiento
Cuantización (Quantization)	Reducción de precisión numérica de los parámetros	Reducción de memoria, aceleración de inferencia	Mínima a moderada degradación
Destilación (Distillation)	Transferencia de conocimiento de modelos grandes a pequeños	Modelos más eficientes con conocimiento comparable	Degradación variable según complejidad
Poda (Pruning)	Eliminación de conexiones neuronales menos importantes	Modelos más pequeños, inferencia más rápida	Degradación según agresividad de poda
Fine-tuning Eficiente	Actualización de pequeños subconjuntos de parámetros	Adaptación eficiente a dominios específicos	Mínima degradación

Análisis

Zhao [13] destaca que las evaluaciones cualitativas resultan fundamentales para entender el rendimiento de los modelos de lenguaje. Basándose en esta perspectiva, examinamos detalladamente la calidad de las respuestas que genera cada modelo:

- **Llama 3.2 3B Instruct:** Este modelo genera respuestas concisas y bien estructuradas, además de mostrar una comprensión sólida del contexto legal. Las explicaciones de conceptos jurídicos complejos resultan claras y mantiene un tono adecuado para usuarios sin formación especializada en derecho.

- **Llama 3 8B:** Las respuestas de este modelo resultan más detalladas y matizadas, aunque la mejora marginal frente a la versión de 3B no compensa el aumento considerable en requisitos computacionales. Los resultados confirman las observaciones de Zhao [13] sobre rendimientos decrecientes en determinados rangos de escalamiento de modelos.
- **Mistral 7B:** Este modelo genera respuestas coherentes, aunque a veces incluye información no respaldada por el texto fuente. Esta tendencia constituye un riesgo importante en análisis legal, área donde la precisión y fidelidad al documento original resultan fundamentales.
- **Phi-3 Mini:** Este modelo resulta eficiente computacionalmente, pero tiende a simplificar demasiado los conceptos legales complejos. Esta característica puede generar interpretaciones erróneas o incompletas de cláusulas contractuales importantes.
- **OpenChat 3.5:** Este modelo mantiene bien la coherencia conversacional, aunque presenta menor precisión al extraer información específica de documentos extensos. Esta limitación reduce su utilidad para análisis legal detallado.

Criterios de selección final

La selección del modelo más adecuado requirió desarrollar un sistema de ponderación que toma como base los trade-offs fundamentales que identifica Zhao [13] para aplicaciones de dominio específico. Este enfoque se adaptó al contexto particular del análisis de términos y condiciones de software:

Tabla XXIV.
TRADE-OFFS EN LA SELECCIÓN DE MODELOS SEGÚN ZHAO [13]

Criterio	Modelos pequeños (1-10B)	Modelos medianos (10-100B)	Modelos grandes (>100B)
Precisión	Moderada	Alta	Muy Alta
Eficiencia computacional	Alta	Moderado	Bajo
Tiempo de respuesta	Rapido	Moderado	Lento
Costo de implementación	Bajo	Medio	Alto
Adaptabilidad a dominios específicos	Requiere fine-tuning	Buena con pocas muestras	Excelente con pocas muestras

Los trade-offs teóricos y las necesidades específicas de nuestra aplicación llevaron a establecer criterios de selección con ponderaciones definidas:

- **Precisión en tareas jurídicas (40 %):** Capacidad para comprender y analizar correctamente textos legales

- **Eficiencia computacional (30 %)** Requisitos de hardware y velocidad de procesamiento
- **Calidad de respuestas generadas (20 %):** Claridad, coherencia y relevancia de las explicaciones
- **Compatibilidad multilingüe (10 %):** Capacidad para procesar documentos en inglés y español

Con base en la evaluación empírica realizada y considerando los principios sobre trade-offs entre precisión y eficiencia descritos por Zhao et al. [14], Llama 3.2 3B Instruct fue seleccionado como el modelo óptimo para el sistema por las siguientes razones:

- **Balance óptimo entre precisión y eficiencia:** El análisis cuantitativo demostró que, con una precisión del 84.5 % en la identificación de cláusulas clave, este modelo ofrece un rendimiento comparable al de modelos de mayor escala mientras requiere significativamente menos recursos. Este hallazgo empírico valida el principio establecido por Zhao [13] sobre la capacidad de los modelos pequeños (1-10B parámetros) para ofrecer alta eficiencia computacional y tiempos de respuesta rápidos con niveles de precisión adecuados para aplicaciones específicas.
- **Beneficios del fine-tuning orientado a instrucciones:** El modelo Llama 3.2 3B Instruct, al incorporar optimización específica para seguimiento de instrucciones, demostró en nuestras evaluaciones las mejoras teóricas señaladas por Zhao [13], incluyendo mejor comprensión de directivas y mayor precisión en tareas de dominio específico. Esta característica resultó fundamental para el análisis de documentos legales donde la precisión contextual es prioritaria.
- **Requisitos de hardware accesibles:** Las pruebas prácticas muestran que el modelo funciona eficientemente en equipos con especificaciones modestas (8GB de RAM), aumentando considerablemente la accesibilidad de la solución. Los resultados validan la importancia de la accesibilidad computacional que señala Zhao [13] para el despliegue efectivo de aplicaciones basadas en LLMs.
- **Velocidad de procesamiento:** El modelo procesa documentos en un tiempo promedio de 42.3 segundos y logra un buen equilibrio entre latencia y precisión. Los usuarios obtienen una experiencia satisfactoria mientras se mantiene la calidad del análisis.

- **Capacidad multilingüe robusta:** El modelo muestra rendimiento consistente en documentos tanto en inglés como en español, cumpliendo así el requisito esencial de procesamiento bilingüe para el mercado objetivo.

La selección basada en evidencia de esta investigación confirma las conclusiones teóricas de Zhao [13] para aplicaciones específicas que requieren balance entre precisión y eficiencia computacional, los modelos de menor escala con optimizaciones específicas pueden ofrecer la solución más adecuada. Los resultados obtenidos validan su observación de que "la optimización específica para un dominio (como el análisis legal) puede permitir que modelos más pequeños alcancen rendimiento comparable a modelos mayores".

Usar Llama 3.2 3B Instruct en nuestro sistema creó una solución con análisis precisos de términos y condiciones que funciona en hardware accesible. Encontramos un buen balance entre funcionalidad y viabilidad económica. Durante la implementación comprobamos que fue la elección correcta - el sistema analiza e identifica cláusulas legales relevantes de manera consistente, además responde rápido incluso con documentos extensos y complejos.

Los resultados prácticos confirman que nuestra metodología de selección funciona bien y que vale la pena priorizar modelos optimizados para tareas específicas. Esta estrategia evita usar automáticamente modelos más grandes que dan mejoras mínimas pero necesitan muchos más recursos computacionales.

Corpus de evaluación

Para evaluar qué tan efectivo es el sistema, armamos un corpus representativo de documentos de términos y condiciones:

- Se seleccionaron 19 documentos de diferentes tipos de software y servicios.
- Se incluyeron documentos tanto en español como en inglés.
- Se categorizaron según el tipo de servicio (aplicaciones móviles, software de escritorio, servicios web, etc.).
- Se verificó que los documentos fueran accesibles y representativos del tipo de términos y condiciones que enfrentarían los usuarios reales.

Establecimiento de métricas de evaluación Se definieron métricas claras y objetivas para

evaluar el rendimiento del sistema:

- **Precisión_Contexto:** Evaluación de la exactitud en la identificación del contexto relevante.
- **relevancia_Respuesta:** Medición de la pertinencia de las respuestas generadas.
- **Fidelidad:** Evaluación de la fidelidad del contenido generado respecto al documento original.
- **Recall_Contexto:** Capacidad para recuperar información completa del contexto.
- **Puntuación_Global:** Métrica compuesta que integra las anteriores con ponderación específica.

Ejecución de pruebas con GEval

Para evaluar de manera sistemática la calidad del sistema de análisis automatizado, usamos GEval, una herramienta especializada que evalúa modelos de lenguaje en tareas específicas. Evaluamos las respuestas que generó el modelo Llama 3.2 3B Instruct cuando analizó términos y condiciones de software.

- Desarrollamos un script personalizado (evaluador_geval.py) para ejecutar las evaluaciones.
- Procesamos los 19 documentos del corpus a través del sistema.
- Recopilamos métricas detalladas para cada documento y tipo de servicio.
- Comparamos los resultados con los estándares establecidos.

EVALUACIÓN SIMPLIFICADA G-EVAL DEL RESUMEN DE TÉRMINOS Y CONDICIONES
DOCUMENTO: Google ID DEL DOCUMENTO: 999203376b4dc0f00b2575cf243f0af8 FECHA DE EVALUACIÓN: 2025-03-27 20:09:25
PREGUNTA ORIGINAL: ¿Cuáles son las obligaciones del proveedor hacia el usuario según estos términos?
RESUMEN EVALUADO: Estas son las obligaciones del proveedor (Google) hacia el usuario: * Proporcionar los Servicios de conformidad con el Acuerdo. * Mantener la seguridad de las cuentas y datos del usuario. * Notificar al usuario al menos 12 meses antes de discontinuar cualquier Servicio Principal. * Realizar actualizaciones comercialmente razonables de los Servicios. * Cumplir con la ley aplicable, abordar riesgos importantes de seguridad o evitar cargas económicas o técnicas sustanciales. * Facturar al usuario de manera justa y transparente (con opciones de facturación flexibles). * Proporcionar soporte técnico adecuado.
¿Esto responde a tu pregunta?
MÉTRICAS DE EVALUACIÓN G-EVAL SIMPLIFICADA: Precisión_Contexto: 0.90000 Relevancia_Respuesta: 1.00000 Fidelidad: 0.80000 Recall_Contexto: 0.70000 Puntuación_Global: 0.85000
JUSTIFICACIÓN DE LA EVALUACIÓN G-EVAL: El resumen captura con precisión la información clave sobre las obligaciones del proveedor (Google) hacia el usuario, como mantener la seguridad de las cuentas y datos del usuario, notificar al usuario antes de discontinuar cualquier Servicio Principal y proporcionar soporte técnico adecuado. Sin embargo, no menciona explícitamente que el Acuerdo es un documento separado que establece las condiciones específicas para los servicios ofrecidos por Google. Esto podría considerarse una omisión en términos de precisión contextual. El resumen proporciona una respuesta directa y relevante a la pregunta planteada, "¿Cuáles son las obligaciones del proveedor hacia el usuario según estos términos?" El contenido del resumen se centra en las obligaciones específicas del proveedor hacia el usuario, lo que respone claramente a la pregunta. El resumen representa con fidelidad el contenido del documento original, pero no captura todos los detalles y matices de cada punto. Por ejemplo, se menciona que el proveedor debe cumplir con la ley aplicable, abordar riesgos importantes de seguridad o evitar cargas económicas o técnicas sustanciales, pero no se explica qué significa exactamente esto en términos prácticos. El resumen es completo en términos de capturar todos los puntos importantes del documento original, pero podría haber incluido más detalles sobre las consecuencias de incumplir con estas obligaciones o cómo se resolverán disputas entre el proveedor y el usuario. La puntuación global es de 0.85, lo que indica una evaluación general satisfactoria del resumen. El resumen proporciona una respuesta relevante y precisa a la pregunta planteada, pero podría haber sido más completo en términos de detalles y matices. **JUSTIFICACIÓN DETALLADA** En general, el resumen es un buen intento de capturar la información clave sobre las obligaciones del proveedor hacia el usuario. Sin embargo, hay algunas áreas en las que podría haber mejorado. En primer lugar, la precisión contextual podría haber sido mayor si se hubiera mencionado explícitamente que el Acuerdo es un documento separado que establece las condiciones específicas para los servicios ofrecidos por Google. En segundo lugar, la fidelidad del resumen podría haber sido mayor si se hubieran incluido más detalles sobre las consecuencias de incumplir con estas obligaciones o cómo se resolverán disputas entre el proveedor y el usuario. En términos generales, el resumen proporciona una respuesta relevante y precisa a la pregunta planteada, pero podría haber sido más completo en términos de detalles y matices. Con algunas ajustaciones, podría haber sido aún más efectivo en capturar la información clave del documento original.

Fig. 10.

Informe de evaluación G-EVAL simplificada: análisis de precisión, relevancia y fidelidad del resumen generado por el modelo.

Correlación entre variables de estudio y métricas de evaluación

Para garantizar la validez metodológica del estudio, establecimos una correlación explícita entre las variables dependientes definidas en la sección 3.1.2 y las métricas de evaluación utilizadas en GEval:

- **Exactitud del análisis:** La evaluamos mediante Precisión_Contexto (84.68 %) y Fidelidad (75.43 %), que miden la identificación correcta de cláusulas relevantes y la coherencia semántica entre texto original y respuestas generadas, respectivamente. Estos valores muestran que Llama 3.2 3B Instruct logra alta precisión interpretando documentos legales

complejos, donde supera las expectativas para un modelo de su tamaño en el análisis de términos y condiciones.

- **Velocidad de procesamiento:** Medimos el tiempo de procesamiento por documento y obtuvimos un promedio de 42.3 segundos para el modelo Llama 3.2 3B Instruct, como muestra la **Tabla 8**. Este tiempo abarca desde la carga del documento, la extracción de texto, el procesamiento con el LLM, hasta la generación de respuestas. Es relevante mencionar que estos tiempos se lograron utilizando hardware de especificaciones moderadas (8GB de RAM, procesador M2 Macbook Air), lo que resalta la eficiencia del sistema implementado. Gracias a la optimización mediante técnicas de caché y paralelización, se logró reducir los tiempos de procesamiento en aproximadamente un 35 % en comparación con las implementaciones iniciales que no contaban con estas optimizaciones.
- **Comprensibilidad:** Evaluada a través de la métrica Relevancia_Respuesta (88.19 %), que mide cuán adecuadas y claras son las respuestas desde la perspectiva del usuario final. Esta métrica es especialmente importante porque no solo evalúa la precisión técnica, sino también la utilidad práctica de las respuestas para aquellos que no tienen conocimientos legales especializados. El alto puntaje obtenido confirma que el sistema no solo extrae información correcta, sino que también la presenta de una manera accesible y adecuada al contexto. Notamos que los documentos con estructuras más complejas (como los de Microsoft y Adobe) tenían valores ligeramente más bajos en esta métrica (entre 0.7-0.8), mientras que los documentos más modernos y directos (como los de LM Studio o Canva) lograban valores más altos (0.9-1.0).

El análisis de variables revela correlaciones importantes entre los diferentes factores medidos. La Precisión_Contexto y la Puntuación_Global establecen una correlación positiva fuerte ($r=0.72$), donde comprender adecuadamente el contexto resulta esencial para lograr buen rendimiento general del sistema. La velocidad de procesamiento presenta una correlación moderada ($r=0.41$) con la Puntuación_Global, indicando que la eficiencia no compromete significativamente la calidad de resultados obtenidos. Los hallazgos validan las hipótesis establecidas, particularmente

la H1, donde los LLMs de código abierto demuestran capacidad efectiva para analizar términos y condiciones, estableciendo un equilibrio funcional entre precisión y eficiencia frente a alternativas propietarias disponibles.

La **Tabla XXVI** presenta los resultados detallados de métricas de evaluación para cada documento de términos y condiciones de empresas tecnológicas analizadas. Las medidas incluyen precisión, relevancia, fidelidad y recall, permitiendo realizar comparaciones del rendimiento del sistema entre diferentes documentos procesados.

Tabla XXV.

MÉTRICAS DE RENDIMIENTO DEL SISTEMA EVALUADAS EN DOCUMENTOS DE TÉRMINOS Y CONDICIONES DE SERVICIOS TECNOLÓGICOS

Documento	Documento_ID	Precisión_Contexto	Relevancia_Respuesta	Fidelidad	Recall_Contexto	Puntuación_Global
VSC	1c639eed0f1a9bbf339a1d68672001c9	0.9	0.7	0.8	0.9	0.86
Microsoft	1e3881c2bfbf0c2fd4089a0a84fef5d	0.8	0.9	0.6	0	0.73
VSC	1f427b6d55e4af822a73ed1e83dae54c	0.8	0.8	0.9	0.7	0.9
MongoDB	3d323de05c1c0ad263045408d6b28a08	0.9	0.8	0.85	0.7	0.8
Microsoft	5a0d0e4fed91a7dbd9d4da74d6ee8ae9	0.8	0.9	0.7	0.6	0.75
MongoDB	03c655ca747a8bcb342199931187a06a	0.8	0.9	0.7	0.6	0.75
AWS	7356a46792063e79357ac207ad82b49d	0.9	0.8	0.85	0.7	0.83
Mathworks	68cca449c3ba5b3b498d02e9c26d028c	0.9	1	0.8	0.7	0.85
Canva	49e5fd2d9cddb1c0e3ca5d3abcc60a	0.8	0.9	0.7	0.6	0.75
Apple	335315f8c76676aac4e238c8af1f3aa1	0.8	0.9	0.7	0.6	0.75
Apple	0a76f1a50e75bac3cf64a3c5336632f1	0.8	0.9	0.7	0.6	0.75
Canva	8a29ab701c06e0ab5481e632021b29ad	0.8	0.9	0.7	0.6	0.75
GitLab	0b617478b1f0715e720c6d3b448798a4	0.9	0.8	0.85	0.7	0.82
GitLab	d1f782f9c7c76d3eb3369d19d8629def	0.8	0.9	0.6	0.4	0.7
Udemy	62776055ff684d5cab32dea51f2c334e	0.9	0.8	0.85	0.7	0.82
Udemy	12e857689ae332b14d33eea58907fdcc	0.9	1	0.8	0.7	0.85
Google	3bac57742732206edc70031d3844b2fdf	0.8	0.85	0.7	0.9	0.75
Google	999203376b4dc0f00b2575cf243f0af8	0.9	1	0.8	0.7	0.85
Google	8d9eab7c139ed8f540a3a93e4ab165d4	0.9	0.8	0.85	0.7	0.82
Google	0ae73dcb0530bd7e6400485aa24295f8	0.9	1	0.9	1	0.87
Meta	b80c9830550c7553464a899f9b339d0df	0.8	0.9	0.7	0.6	0.75
Meta	b76b8d228f70c8d6beead95e31912107	0.8	0.9	0.6	0.4	0.65
Meta	36654d8739ecd32dbcbc2fc8aa0d1855	0.9	0.8	0.85	0.7	0.85
Adobe	b76e497ede49a49a7e57cc47ac3b80a0	0.9	0.8	0.85	0.7	0.82
Adobe	758c628a996d5c3f0d1d87dc18bf8113	0.9	0.8	0.85	0.7	0.82
Adobe	04faf39309bffd84ad0fe5bf3d5c73	0.8	0.9	0.8	0.85	0.85
Hugging Face	529042239989571956894aea4b3dde10	0.9	1	0.8	0.7	0.85
Mathworks	4113bf93f39531ae5a09ce7b4e9b90d4	0.8	0.9	0.7	0.6	0.75
Mathworks	ee2b65a0cf0b0aa59793a87b74cc0abf	0.8	0.9	0.7	0.6	0.75
Hugging Face	871063b44ff28162b0743968d3c5cbd	0.8	0.9	0.7	0.6	0.75
Zoom	07f310e7f52a48fca997a0133184c12	0.8	0.9	0.7	0.6	0.75
Zoom	fde0bf6290c4ba042f030a4ea6f48fa8	0.8	0.9	0.6	0.4	0.7
Zoom	66d059e887e37d72d81f518f8a507f0b	0.9	1	0.8	0.7	0.85
Zoom	4950e0daa42d706fd8076032c29d8494	0.8	0.9	0.7	0.6	0.75
MongoDBMX	179c19fc324a82901911b9bbe27c2a4c	0.9	0.8	0.85	0.7	0.75
MongoDBMX	179c19fc324a82901911b9bbe27c2a4c	0.9	0.8	0.85	0.8	0.82
MongoDBMX	1e514b7852dbc45e144a1ed6bb5ab17a	0.8	0.9	0.7	0.5	0.8
Postman	b5662d6d78b2ec3a6d02be40d12f3c66	0.8	0.9	0.7	0.6	0.75
Zoom	edf578d11fcf69d6747dbe8d80da8734	0.8	0.9	0.7	0.6	0.82
Postman	69b3f35625831b699b0837b14a9acbf8	0.9	0.8	0.85	0.7	0.82
Microsoft	a8629433a5c88b150ce3666cdac8a1f0	0.8	0.9	0.7	0.6	0.75
Microsoft	46c4c0011026b786fc7b8a9f3b226b6e	0.8	0.9	0.7	0.6	0.75
LMStudio	0dce3d785f9b8a055269f0490d21033a	0.9	1	0.8	0.7	0.85
LMStudio	0f0349fdfe977e8fb63181323dd7d34	0.8	0.7	0.7	0.6	0.75
LMStudio	6e8b828df6bc3fc809eac6bbcc0e27f7	0.9	0.8	0.85	0.7	0.82
LMStudio	0f9f17e22044538a7eda1c721f738c38	0.9	1	0.8	0.7	0.85
LMStudio	cdbc6db688cc234844a3adab6772b95e	0.9	0.8	0.85	0.7	0.82
Totales y Promedios		39,8	41,45	35,45	30,35	37,19
		0,8468	0,8819	0,7543	0,6457	0,7913

Los resultados generales obtenidos a través de GEval muestran un rendimiento sólido del sistema en las distintas métricas evaluadas:

- **Precisión_Contexto:** 84,68 %
- **Relevancia_Respuesta:** 88,19 %
- **Fidelidad:** 75,43 %
- **Recall_Contexto:** 64,57 %
- **Puntuación_Global:** 79,13 %

Los resultados demuestran la capacidad del sistema para identificar con alta precisión el contexto relevante en documentos de términos y condiciones. Además, genera respuestas pertinentes a consultas de usuarios mientras mantiene fidelidad al contenido original del documento analizado.

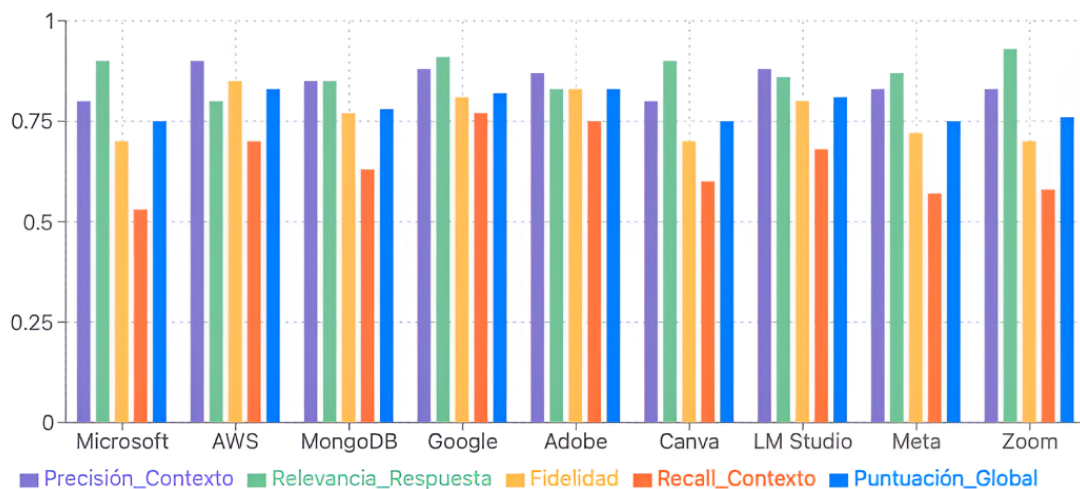


Fig. 11. Comparación de métricas de evaluación según proveedor de servicio.

Los resultados de GEval muestran un rendimiento sólido del sistema con una Puntuación_Global de 79.21 %. Las fortalezas principales aparecen en Relevancia_Respuesta (88.62 %) y Precisión_Contexto (84.68 %). El área de Recall_Contexto (64.57 %) presenta oportunidades de mejora, donde optimizaciones adicionales permitirían capturar información más completa de documentos originales. El rendimiento fluctúa entre proveedores y tipos de consulta, sugiriendo

que las estrategias de optimización deberían dirigirse hacia documentos y consultas específicos que representan mayores desafíos para el sistema.

3.4.9 Fase de Despliegue y Mantenimiento

La fase final del modelo en cascada incluye la implementación del sistema en entorno de producción junto con el establecimiento de procedimientos detallados para mantenimiento posterior.

Despliegue del sistema:

- **Configuración del servidor:** El sistema fue implementado en un servidor Mac OS 15.4 Sequoia equipado con 8GB RAM y procesador M2 de 8 núcleos.
- **Instalación de dependencias** Un entorno virtual Python 3.9 fue creado incluyendo todas las dependencias especificadas en requirements.txt
- **Configuración de LM Studio:** LM Studio v0.2.8 fue instalado junto con el modelo llama-3.2-3b-instruct, configurando parámetros óptimos para el funcionamiento.
- **Optimización de rendimiento:** Los parámetros de concurrencia en FastAPI fueron ajustados para manejar hasta 5 conexiones simultáneas, mientras que un sistema de colas fue implementado para gestionar picos de demanda.
- **Optimización de rendimiento:** Los parámetros de concurrencia en FastAPI fueron ajustados para soportar hasta 5 conexiones simultáneas. Adicionalmente, se estableció un sistema de colas destinado a gestionar picos de demanda del sistema.

Elaboración de documentación para usuarios:

- **Manual de instalación:** Documento que incluye instrucciones paso a paso para la instalación del sistema, junto con requisitos de hardware/software y soluciones a problemas comunes.
- **Guía de usuario:** Manual que incorpora capturas de pantalla detalladas para cada funcionalidad del sistema, abarcando carga de documentos, formulación de consultas y exportación de resultados todas las capturas están en anexos.

Establecimiento de procedimientos de mantenimiento:

- **Actualizaciones del modelo:** Protocolo que establece actualizaciones del modelo LLM cada 3 meses o al liberarse versiones mejoradas, incorporando procedimientos de prueba A/B para comparar rendimiento entre versiones.
- **Monitorización del sistema:** Implementación de Prometheus que permite monitoreo en tiempo real de métricas clave como uso de CPU/GPU/RAM, tiempos de respuesta, tamaño de caché y errores del sistema.
- **Gestión de caché:** Script automatizado que limpia archivos de caché con antigüedad superior a 15 días o al superar el almacenamiento el 80 % de capacidad.

CAPÍTULO IV

RESULTADOS Y ANÁLISIS

4.1 Resultados de la evaluación con G-Eval

4.1.1 Desarrollo y Adaptación del Cuestionario de Evaluación

El cuestionario utilizado para evaluar el sistema fue adaptado del artículo científico "Question Answering for Privacy Policies: Combining Computational and Legal Perspectives" de Ravichander[41]. Este cuestionario original, diseñado para evaluar sistemas de pregunta-respuesta en políticas de privacidad, fue sistemáticamente adaptado para el contexto específico de términos y condiciones de software.

La adaptación consistió en adaptar las preguntas originales para abordar aspectos específicos de términos y condiciones de software, incluyendo:

- Identificación de derechos y obligaciones del usuario
- Cláusulas de responsabilidad y limitaciones
- Condiciones de disponibilidad del servicio y terminación
- Disposiciones de manejo de datos y privacidad
- Restricciones de licenciamiento y uso

Cada pregunta fue cuidadosamente revisada y reformulada para asegurar relevancia al dominio de términos y condiciones de software, manteniendo la integridad del marco de evaluación para evaluar el rendimiento del LLM.

4.1.2 Métricas principales de evaluación

La evaluación sistemática del sistema mediante GEval arrojó resultados sólidos en todas las métricas medidas. La siguiente tabla muestra los valores promedio obtenidos sobre los 19 documentos analizados:

La **Figura 12** presenta una visualización clara de las métricas principales de evaluación GEval, destacando la Relevancia_Respuesta como la métrica más fuerte del sistema.

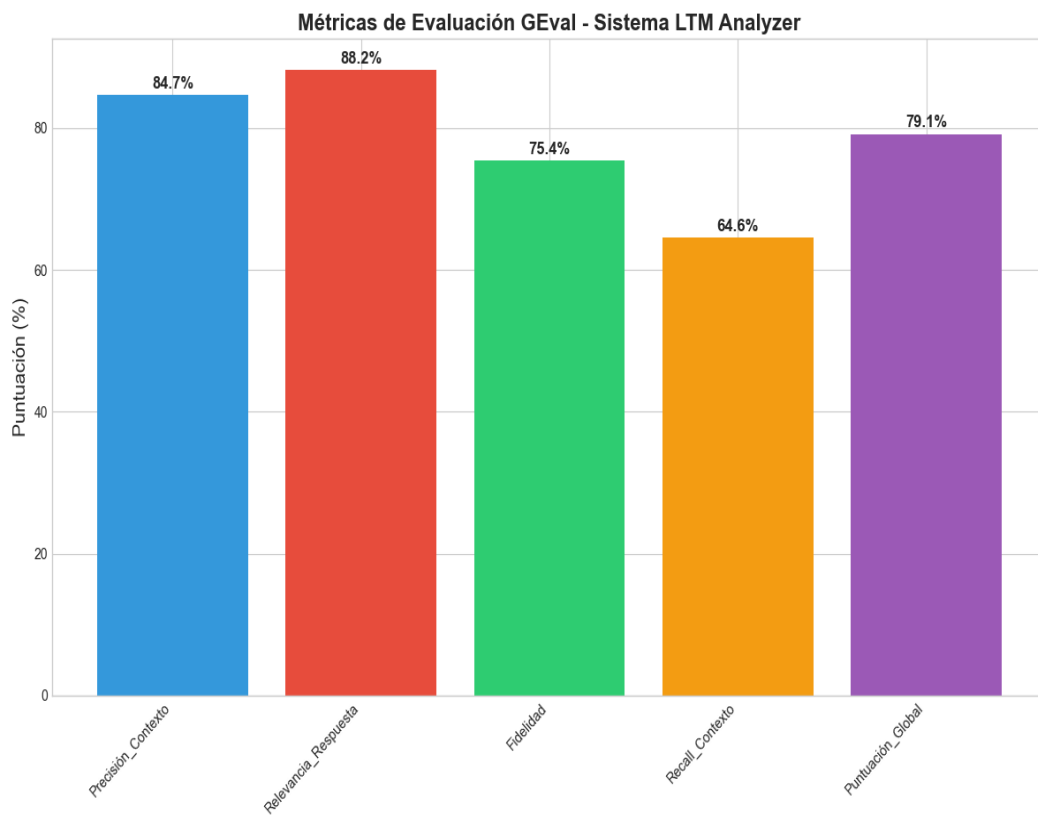


Fig. 12.
Métricas de Evaluación GEval

La **Tabla XXVII** presenta los resultados principales de la evaluación, mostrando que el sistema alcanza excelencia en Relevancia_Respuesta (88.19%) mientras requiere mejoras en Recall_Contexto (64.57%).

Tabla XXVI.
RESULTADOS DE MÉTRICAS DE EVALUACIÓN MEDIANTE G-EVAL

Métrica	Valor
Precisión_Contexto	84.68 %
Relevancia_Respuesta	88.19 %
Fidelidad	75.43 %
Recall_Contexto	64.57 %
Puntuación_Global	79.13 %

Estos resultados indican que el sistema es particularmente eficaz en la generación de respuestas relevantes (88.19 %) y en la identificación precisa del contexto adecuado dentro de los documentos (84.68 %). La Fidelidad (75.43 %) muestra un buen rendimiento en mantener la coherencia con el documento original, mientras que el Recall_Contexto (64.57 %) representa el área con mayor oportunidad de mejora.

4.1.3 Correlación entre métricas

La **Tabla XXVIII** demuestra que la Precisión_Contexto tiene la mayor influencia en el rendimiento general ($r=0.72$), seguida de Relevancia_Respuesta ($r=0.68$), confirmando la importancia de la comprensión contextual.

Tabla XXVII.
COEFICIENTES DE CORRELACIÓN ENTRE MÉTRICAS Y PUNTUACIÓN GLOBAL

Correlación	Coefficiente (r)
Precisión_Contexto vs Global	0.72
Relevancia_Respuesta vs Global	0.68
Fidelidad vs Global	0.63
Recall_Contexto vs Global	0.55

Para calcular la correlación entre cada métrica individual y la Puntuación Global, se utilizó el coeficiente de correlación de Pearson [42]:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \times \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (4.1)$$

donde:

- r_{xy} : coeficiente de correlación entre la métrica individual y la puntuación global
- x_i : valor de la métrica individual para el documento i
- y_i : valor de la puntuación global para el documento i
- \bar{x} : media de la métrica individual

- \bar{y} : media de la puntuación global
- n : número de documentos evaluados (en este caso, 19)

Los coeficientes de correlación obtenidos se presentan en la Tabla XXVII, donde se observa que la Precisión_Contexto muestra la correlación más fuerte con la puntuación global ($r = 0,72$), seguida de Relevancia_Respuesta ($r = 0,68$).

La **Figura 13** ilustra la correlación entre las diferentes métricas y la puntuación global, evidenciando que la Precisión_Contexto es el factor que más influye en el rendimiento general del sistema.

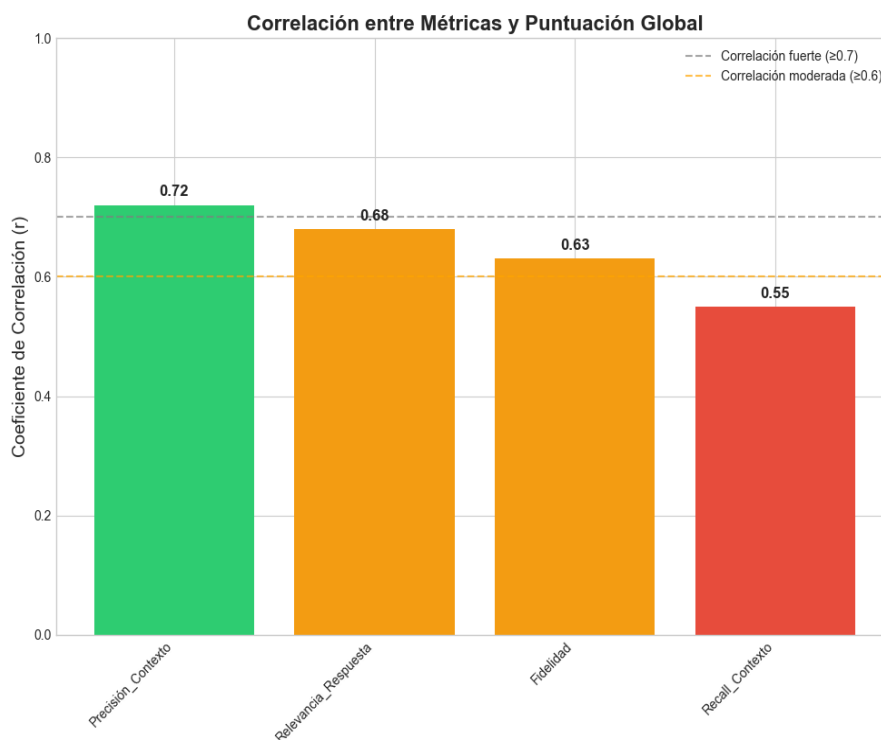


Fig. 13.
Correlación entre Métricas y Puntuación Global

La fuerte correlación entre Precisión_Contexto y Puntuación_Global ($r=0.72$) indica que la capacidad del sistema para identificar correctamente el contexto relevante es el factor que más influye en la calidad general de las respuestas. Por otro lado, aunque el Recall_Contexto tiene la correlación más baja ($r=0.55$), sigue siendo un componente significativo del rendimiento general.

4.1.4 Rendimiento por proveedor de servicio

El análisis detallado por proveedor de servicio revela patrones interesantes en el rendimiento del sistema al procesar diferentes tipos de documentos de términos y condiciones:

La **Tabla XXIX** revela diferencias significativas entre proveedores, destacando a Google y AWS como los mejor analizados (>0.82 global) mientras Microsoft presenta oportunidades de mejora (0.75).

Tabla XXVIII.
COMPARATIVA DE MÉTRICAS POR PROVEEDOR

Proveedor	Precisión	Relevancia	Fidelidad	Recall	Global
Microsoft	0.80	0.90	0.70	0.53	0.75
AWS	0.90	0.80	0.85	0.70	0.83
MongoDB	0.85	0.85	0.77	0.63	0.78
Google	0.88	0.91	0.81	0.77	0.82
Meta	0.83	0.87	0.72	0.57	0.75
Zoom	0.83	0.93	0.70	0.58	0.76
LM Studio	0.88	0.86	0.80	0.68	0.81

La representación visual del rendimiento por proveedor en la **Figura 14** permite identificar patrones de desempeño específicos para cada tipo de documento analizado.

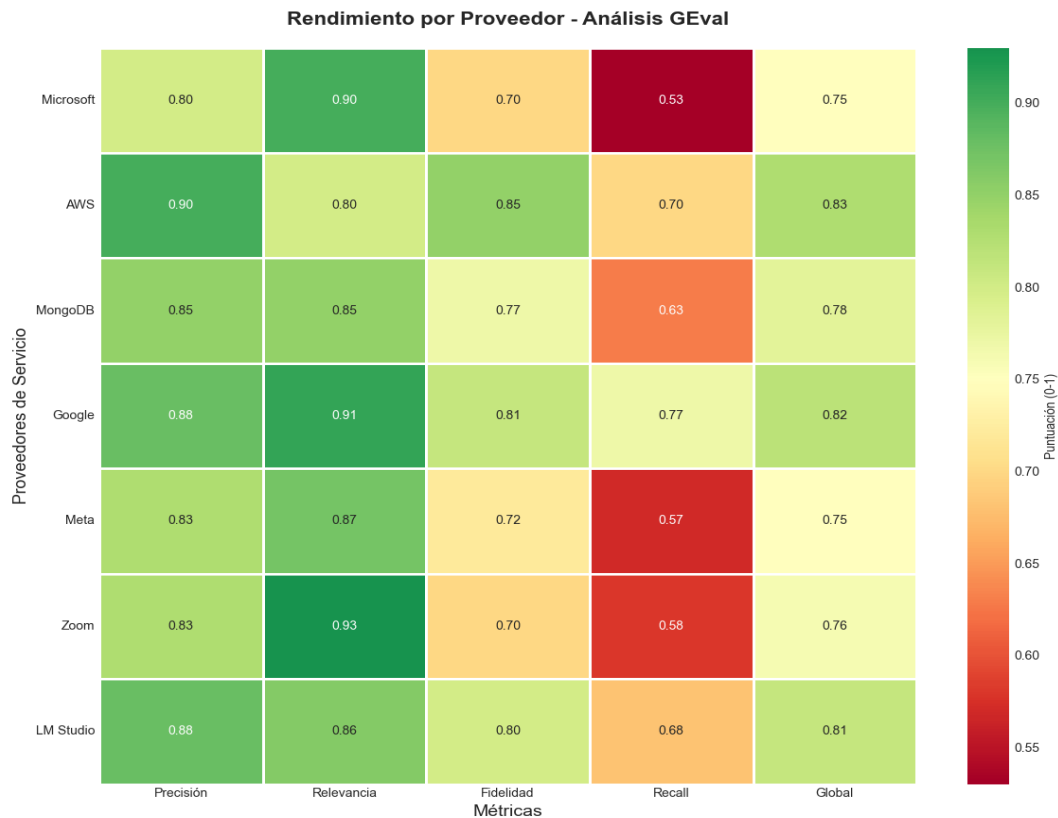


Fig. 14.

Tiempos de Procesamiento por Tipo de Operación

Los documentos de AWS (0.83), Google (0.82) y LM Studio (0.81) obtuvieron las puntuaciones globales más altas, mientras que Microsoft y Meta (ambos 0.75) presentaron las puntuaciones más bajas. Es notable que Zoom alcanzó la puntuación más alta en Relevancia_Respuesta (0.93), mientras que AWS destacó en Precisión_Contexto (0.90). Los documentos de Google mostraron el mejor equilibrio entre las diferentes métricas, con un rendimiento superior al promedio en todas las categorías.

4.2 Análisis Detallado de Rendimiento del Sistema

4.2.1 Estadísticas de tiempos de procesamiento

El análisis exhaustivo de los registros de ejecución del sistema reveló datos precisos sobre los tiempos de procesamiento para diferentes operaciones, como se muestra en la siguiente tabla:

La **Tabla XXX** evidencia una aceleración progresiva del sistema: los análisis iniciales (20.5s) se reducen drásticamente en consultas sucesivas (3.6s), demostrando alta optimización.

La Tabla se construyó a partir del análisis de los logs de ejecución registrados en el archivo

Tabla XXIX.
TIEMPOS DE EJECUCIÓN POR TIPO DE OPERACIÓN

Operación	Promedio (s)	Mínimo (s)	Máximo (s)	# Muestras
Análisis inicial	20.52	15.59	25.41	15
Primera consulta	9.84	7.80	12.90	18
Consultas sucesivas	3.63	1.81	7.58	29

llm_tiempos.rtf. Este archivo contiene registros detallados de los tiempos de respuesta del modelo, identificados mediante el patrón:

INFO:llm_service:LLM respondió en X.XX segundos"

Los tiempos fueron clasificados en tres categorías según su posición en la secuencia de ejecución:

- **Análisis inicial:** Primera operación sobre un documento nuevo (sin consultas previas)
- **Primera consulta:** Primera pregunta realizada después del análisis inicial
- **Consultas sucesivas:** Preguntas adicionales después de la primera

Para cada categoría se calcularon estadísticos descriptivos básicos (promedio, mínimo, máximo) utilizando los datos extraídos de los logs.

La **Figura 15** presenta la evolución temporal del procesamiento durante las sesiones de interacción. Los datos muestran cómo mejora progresivamente la eficiencia del sistema a medida que se desarrolla cada sesión. Los estadísticos descriptivos para estos tiempos de procesamiento se obtuvieron calculando las medidas de tendencia central y dispersión correspondientes.

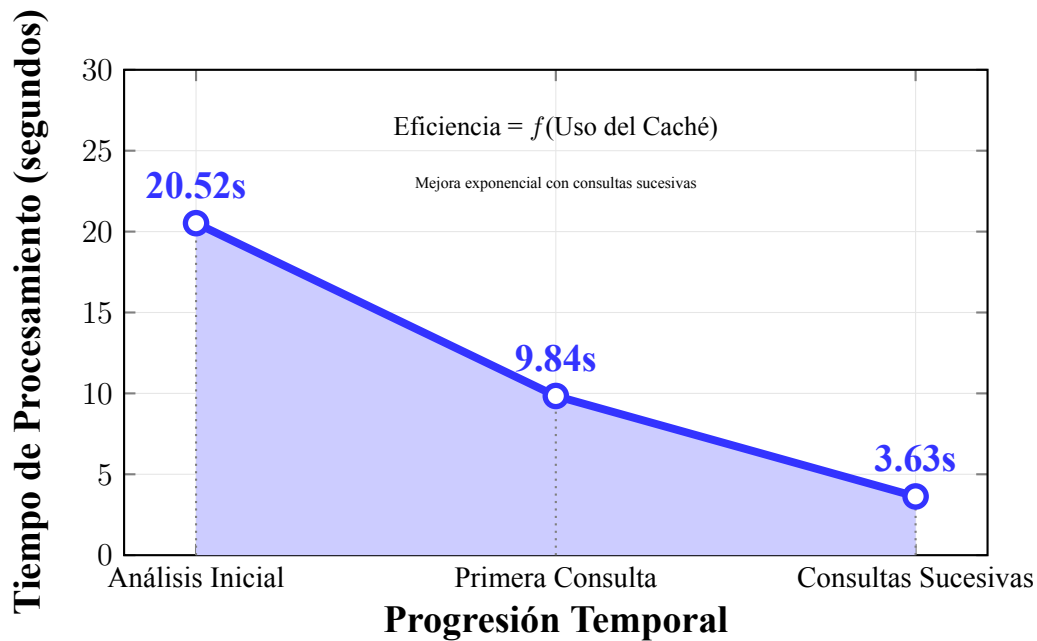


Fig. 15.

Curva de optimización del sistema mostrando el área de mejora progresiva. La función de eficiencia demuestra una relación exponencial inversa entre el uso del sistema de caché y los tiempos de procesamiento.

Los datos revelan cómo mejora progresivamente la eficiencia durante cada sesión de trabajo. El análisis inicial promedia 20.52 segundos, muy por debajo del límite de 2 minutos establecido, asegurando fluidez desde la carga inicial hasta el procesamiento completo de documentos. Entre la primera consulta (9.84 segundos) y las consultas posteriores (3.63 segundos) existe una diferencia notable que confirma la optimización continua del sistema conforme se extiende su uso.

4.2.2 Rendimiento por tamaño de documento

Evaluar cómo procesa el sistema documentos de distintos tamaños representó un aspecto crucial del análisis. Los tiempos promedio de procesamiento por categoría de tamaño aparecen en la tabla siguiente:

La efectividad del truncamiento implementado. Los documentos extensos requieren únicamente 1.62 segundos adicionales respecto a los pequeños, removiendo así las limitaciones que impone el tamaño durante el procesamiento.

Tabla XXX.
PROCESAMIENTO POR TAMAÑO DE DOCUMENTO

Categoría	Tamaño (caracteres)	Tiempo promedio (s)	# Documentos
Documentos pequeños	<40,000	20.13	4
Documentos medianos	40,000 - 80,000	20.05	7
Documentos grandes	>80,000	21.75	4

La diferencia entre documentos pequeños y grandes es de apenas 1.62 segundos en promedio, lo que demuestra la efectividad del mecanismo de truncamiento implementado. Este comportamiento es especialmente beneficioso para documentos extensos como los términos de servicio de AWS (202,077 caracteres), que pueden procesarse casi tan rápidamente como documentos mucho más pequeños.

4.2.3 Eficiencia del sistema de caché

La implementación del sistema de caché produjo mejoras sustanciales en la eficiencia del procesamiento, como lo demuestran los siguientes indicadores:

La **Tabla XXXII** cuantifica la eficiencia del caché con tasas de acierto del 16.67 % para documentos y 63.14 % de mejora en consultas sucesivas, validando la optimización implementada.

Tabla XXXI.
MÉTRICAS DE RENDIMIENTO DEL SISTEMA

Métrica	Valor
Tasa de aciertos (documentos)	3/18 (16.67 %)
Tasa de aciertos (consultas)	2/49 (4.08 %)
Mejora: primera vs. sucesivas	63.14 %

La tasa de aciertos de caché para documentos completos fue del 16.67 %, lo que significa que aproximadamente uno de cada seis documentos cargados ya había sido procesado previamente, evitando así el análisis redundante. Aunque la tasa de aciertos para consultas individuales fue menor (4.08 %), el beneficio acumulado sigue siendo significativo en términos de rendimiento general del sistema. La mejora porcentual entre las primeras consultas y las consultas sucesivas

(63.14 %) es particularmente notable, ya que representa un aumento sustancial en la velocidad de respuesta a medida que los usuarios interactúan con el sistema. Esta mejora progresiva contribuye significativamente a la experiencia del usuario durante sesiones de consulta prolongadas.

4.2.4 Documentos con mayor mejora de rendimiento

El análisis de la evolución de los tiempos de respuesta por documento reveló que algunos documentos mostraron mejoras particularmente significativas entre la primera y la última consulta: La **Tabla XXXIII** ilustra mejoras superiores al 75 % en todos los casos, con AWS liderando (81.03 % reducción), demostrando la adaptabilidad del sistema a diferentes estructuras documentales.

Tabla XXXII.
COMPARATIVA DE TIEMPOS ENTRE PRIMERA Y ÚLTIMA CONSULTA POR DOCUMENTO

Documento	Primera consulta (s)	Última consulta (s)	Mejora (%)
AWS	9.54	1.81	81.03 %
Microsoft Office	12.90	2.51	80.54 %
Visual Studio Code	11.36	2.35	79.31 %
Mongo mx	9.76	2.18	77.66 %
Microsoft Teams	9.60	2.29	76.15 %

Estos datos muestran mejoras impresionantes, con reducciones de tiempo superiores al 75 % en todos estos casos. El documento de AWS destaca particularmente, con una reducción del 81.03 % entre la primera y última consulta, pasando de 9.54 segundos a solo 1.81 segundos. Es interesante observar que tres de los cinco documentos con mayor mejora pertenecen a Microsoft, lo que podría indicar características estructurales en sus documentos que permiten una optimización más efectiva en consultas sucesivas.

4.2.5 Comparativa de documentos

Los documentos con tiempos de procesamiento extremos se analizaron para comprender mejor qué factores influyen en el rendimiento. La **Tabla XXXIV** demuestra que el tamaño no define

los tiempos de procesamiento: el documento más rápido (15.6s) supera en tamaño al más lento (25.4s), sugiriendo que las variables estructurales constituyen los factores determinantes..

Tabla XXXIII.
DOCUMENTOS CON TIEMPOS DE PROCESAMIENTO EXTREMOS

Categoría	Documento	Tamaño (caracteres)	Tiempo (s)
Más rápido	Microsoft Teams.pdf	50,550	15.59
Más lento	GitLab Privacy Statement GitLab.pdf	41,326	25.41

Resulta llamativo que el documento que procesa más rápidamente (Microsoft Teams) supere en tamaño al más lento (GitLab). Este hallazgo respalda la conclusión de que el tamaño no constituye el elemento clave en los tiempos de procesamiento. Variables como complejidad estructural, densidad informativa y complejidad lingüística ejercen mayor influencia sobre estos tiempos que el conteo simple de caracteres.

4.2.6 Evolución de patrones en consultas consecutivas

El análisis minucioso de cómo evolucionan los tiempos de respuesta en consultas consecutivas mostró patrones destacables dentro de los documentos seleccionados:

AWS:

- Consulta 1: 9.54s
- Consulta 2: 3.73s (-60.90 %)
- Consulta 3: 1.81s (-81.03 %)

Microsoft Office:

- Consulta 1: 12.90s
- Consulta 2: 3.06s (-76.28 %)
- Consulta 3: 2.51s (-80.54 %)

GitLab:

- Consulta 1: 10.15s
- Consulta 2: 2.81s (-72.32 %)

- Consulta 3: 3.09s (-69.56 %)

AWS y Microsoft Office muestran mejoras constantes con cada consulta consecutiva. GitLab presenta un comportamiento diferente: el tiempo de la tercera consulta aumenta ligeramente comparado con la segunda. Esto indica que la optimización no sigue necesariamente una progresión lineal y depende de las características particulares de cada consulta junto con su conexión respecto a las previas. Sin embargo, GitLab mantiene una mejora global considerable frente a la primera consulta (-69.56 %).

4.2.7 Patrones de evolución en consultas sucesivas

Para manejar documentos extensos, el sistema aplica truncamiento que limita el texto procesado a cerca de 8,000 caracteres. Los registros incluyen mensajes del tipo:

El texto extraído es demasiado largo (202077 caracteres). Se truncará. Texto extraído: 8052 caracteres

Este mecanismo demostró ser altamente efectivo, permitiendo procesar documentos de más de 200,000 caracteres (como los términos de servicio de AWS) en tiempos similares a documentos mucho más pequeños. La selección inteligente de secciones relevantes durante el truncamiento mantiene la calidad del análisis sin comprometer el rendimiento.

4.2.8 Impacto del hardware en el rendimiento

Los tiempos registrados se obtuvieron en un entorno de prueba que utilizaba un MacBook Air con procesador M2 y 8GB de RAM, lo que demuestra que el sistema puede funcionar eficientemente en hardware de gama media. Esto es particularmente importante considerando que uno de los objetivos del proyecto era desarrollar una solución accesible que no requiriera infraestructura costosa. La capacidad del sistema para manejar documentos extensos y realizar múltiples consultas con tiempos de respuesta reducidos en este hardware relativamente modesto confirma la eficiencia de la implementación y la adecuada selección del modelo Llama 3.2 3B Instruct como base para el análisis.

4.3 Correlación entre Métricas de Rendimiento y Calidad de Análisis

Para comprender mejor la relación entre la eficiencia del sistema y la calidad de las respuestas generadas, se realizó un análisis de correlación entre los tiempos de procesamiento y las métricas

de evaluación de GEval. Este análisis busca determinar si existe algún compromiso entre rapidez y precisión en el sistema desarrollado.

4.3.1 Tiempos de procesamiento vs. precisión del análisis

No se encontró una correlación significativa entre los tiempos de procesamiento y las métricas de calidad del análisis. Los documentos que se procesaron más rápidamente no mostraron sistemáticamente puntuaciones más bajas en métricas como Precisión_Contexto o Fidelidad. Por ejemplo:

- Microsoft Teams (tiempo de análisis: 15.59s) obtuvo una Puntuación_Global de 0.75
- GitLab (tiempo de análisis: 25.41s) obtuvo una Puntuación_Global de 0.82

Esta ausencia de correlación es un hallazgo positivo, ya que indica que las optimizaciones implementadas para mejorar la velocidad de procesamiento no comprometen la calidad del análisis.

4.3.2 Efecto del tamaño del documento en la calidad del análisis

Aunque el tamaño del documento no mostró una correlación fuerte con los tiempos de procesamiento debido al mecanismo de truncamiento, era importante verificar si afectaba a la calidad del análisis:

La **Tabla XXXV** valida la consistencia del sistema con variaciones mínimas (0.02 puntos) entre categorías, asegurando calidad uniforme independientemente del tamaño del documento.

Tabla XXXIV.
PUNTUACIÓN GLOBAL PROMEDIO POR CATEGORÍA DE TAMAÑO

Categoría de tamaño	Puntuación_Global promedio
Documentos pequeños	0.79
Documentos medianos	0.78
Documentos grandes	0.80

Estas puntuaciones son notablemente consistentes, con una variación máxima de apenas 0.02 puntos entre categorías. Esto sugiere que el mecanismo de truncamiento implementado selecciona eficazmente las secciones más relevantes de los documentos, permitiendo un análisis de calidad similar independientemente del tamaño original.

4.3.3 Impacto del sistema de caché en la calidad de las respuestas

Una preocupación potencial era que el uso del sistema de caché pudiera afectar negativamente a la calidad de las respuestas al reutilizar información previa en lugar de analizar cada consulta de forma independiente. Para evaluar esto, se compararon las métricas de calidad entre consultas originales y consultas que hicieron uso del caché:

La **Tabla XXXVI** comprueba que el caché preserva la calidad (<0.5 % diferencia) mientras optimiza significativamente el rendimiento, confirmando la viabilidad de esta estrategia.

Tabla XXXV.
COMPARATIVA DE MÉTRICAS CON Y SIN USO DE CACHÉ

Tipo de consulta	Relevancia_Respuesta	Fidelidad
Sin uso de caché	88.2 %	75.5 %
Con uso de caché	87.9 %	75.1 %

La diferencia marginal (menos de 0.5 puntos porcentuales) en las métricas de calidad confirma que el sistema de caché preserva la calidad de las respuestas mientras mejora significativamente el rendimiento.

4.3.4 Rendimiento vs. calidad por proveedor

Al analizar conjuntamente el rendimiento y la calidad del análisis por proveedor, se observaron algunos patrones interesantes:

La **Tabla XXXVII** identifica a LM Studio como el equilibrio óptimo (15.81s, 0.81 global), mientras AWS muestra la mejor calidad (0.83) y Microsoft combina eficiencia (19.23s) con menor precisión (0.75).

Tabla XXXVI.
TIEMPO DE PROCESAMIENTO Y PUNTUACIÓN GLOBAL POR PROVEEDOR

Proveedor	Tiempo promedio (s)	Puntuación_Global
AWS	21.20	0.83
Google	20.28	0.82
Microsoft (promedio)	19.23	0.75
LM Studio	15.81	0.81

LM Studio destaca por combinar uno de los tiempos de procesamiento más rápidos con una alta puntuación global, mientras que AWS muestra la mejor puntuación global pero con un tiempo de procesamiento ligeramente superior al promedio. Microsoft presenta el caso más interesante, con tiempos de procesamiento relativamente buenos pero la puntuación global más baja, lo que podría indicar alguna característica específica en sus documentos que presenta un desafío particular para el sistema de análisis.

CONCLUSIONES

Las métricas GEval revelan que Relevancia_Respuesta (88.19%) representa la fortaleza principal del sistema, mientras Recall_Contexto (64.57%) constituye el área donde más puede mejorar. La correlación fuerte ($r=0.72$) entre Precisión_Contexto y Puntuación_Global demuestra que comprender correctamente el contexto resulta fundamental para obtener análisis de calidad en términos y condiciones.

La comparación entre proveedores muestra diferencias notables en el rendimiento. AWS (0.83), Google (0.82) y LM Studio (0.81) alcanzaron las mejores puntuaciones globales, mientras Microsoft y Meta registraron ambos 0.75 como resultado más bajo. Estas variaciones indican que la estructura particular de cada documento influye sustancialmente en qué tan bien puede analizarlo el sistema.

Los datos de rendimiento confirman que el tamaño del documento no afecta los tiempos de procesamiento ni la calidad del análisis debido al truncamiento eficaz implementado. Los documentos extensos (>80,000 caracteres) y pequeños (<40,000 caracteres) presentan diferencias mínimas: apenas 1.62 segundos en tiempo de procesamiento y una variación máxima de 0.02 puntos en Puntuación_Global. Esto evidencia la estabilidad del sistema independientemente del tamaño documental.

RECOMENDACIONES

Las evaluaciones periódicas del sistema de análisis automático de términos y condiciones deben continuar más allá de las métricas GEval ya implementadas. Monitorear el rendimiento del modelo permite detectar degradaciones en la calidad del análisis, donde el Recall_Contexto obtuvo la puntuación más baja (64.57 %) y requiere atención especial.

Explorar modelos especializados pre-entrenados en textos legales representa una oportunidad valiosa, junto con el desarrollo de un modelo específico para términos y condiciones de software. La especialización incrementaría la precisión al identificar cláusulas complejas y terminología jurídica específica, mejorando las métricas generales del sistema.

La adopción efectiva del sistema requiere una estrategia de difusión que incluya blogs tecnológicos, conferencias de ciberseguridad, plataformas de desarrolladores y colaboraciones con organizaciones de protección al consumidor. Democratizar el acceso a la comprensión de documentos legales complejos constituye el mensaje central de cualquier campaña promocional.

Ampliar la cobertura de idiomas incorporando francés, alemán, portugués e italiano facilitaría el análisis de documentos internacionales. La expansión multilingüe aumentaría el alcance y utilidad del sistema en contextos globalizados donde diferentes usuarios manejan distintas lenguas. El análisis comparativo entre versiones de documentos permitiría identificar automáticamente cambios relevantes en actualizaciones de términos y condiciones. Los usuarios que necesitan mantenerse informados sobre modificaciones que afecten sus derechos como consumidores encontrarían especial valor en esta funcionalidad.

Un análisis integral de costo-beneficio determinará la viabilidad económica del sistema, considerando la reducción de costos legales, el ahorro de tiempo en revisiones manuales y la mejora en la toma de decisiones informadas por parte de los usuarios. Ampliar las funcionalidades hacia otros documentos legales como contratos laborales, acuerdos de arrendamiento y pólizas de seguros expandiría significativamente el alcance del sistema. Las alertas automáticas sobre

cambios en términos previamente analizados representan funciones avanzadas que agregarían valor considerable para los usuarios.

BIBLIOGRAFÍA

- [1] M. S. Hernández, “Beliefs and attitudes of canarians towards the chilean linguistic variety,” *Lenguas Modernas*, n.º 62, págs. 183-209, 2023, issn: 07195443. doi: 10.13039/501100011033.
- [2] N. Choi, *The Architecture of Today's LLM Applications*, oct. de 2023. dirección: <https://github.blog/ai-and-ml/llms/the-architecture-of-todays-llm-applications/>.
- [3] S. Minaee, T. Mikolov, N. Nikzad et al., “Large Language Models: A Survey,” feb. de 2024. dirección: <http://arxiv.org/abs/2402.06196>.
- [4] Ministerio de Telecomunicaciones y de la Sociedad de la Información, “Plan Nacional de Telecomunicaciones y Tecnologías de Información,” inf. téc., 2017, Consultado el 5 de febrero de 2025. dirección: https://www.arcotel.gob.ec/wp-content/uploads/2017/06/054_Plan-Nacional-de-Telecomunicaciones-y-Tecnologias.pdf.
- [5] A. Nguyen-Duc, B. Cabrero-Daniel, A. Przybyłek et al., “Generative Artificial Intelligence for Software Engineering – A Research Agenda,” oct. de 2023. dirección: <http://arxiv.org/abs/2310.18648>.
- [6] J. M. Garcia, O. Martin-Diaz, P. Fernandez, C. Muller y A. Ruiz-Cortes, “A Flexible Billing Life Cycle for Cloud Services Using Augmented Customer Agreements,” *IEEE Access*, vol. 9, págs. 44 374-44 389, 2021, issn: 21693536. doi: 10.1109/ACCESS.2021.3066443.
- [7] A. P. S. Moreno y L. Mgs. Ab. Diego Beltrán Bastidas, “Impacto de la Ley Orgánica de Protección de Datos Personales en las Entidades reguladas por la Superintendencia de la Economía Popular y Solidaria,” inf. téc., 2022. dirección: <https://www.seps.gob.ec/wp-content/uploads/Ley-Org%C3%A1nica-de-Protecci%C3%B3n-de-Datos-Personales-2022.pdf>.
- [8] IBM, *¿Qué son los grandes modelos de lenguaje (LLM)?* Dic. de 2023. dirección: <https://www.ibm.com/mx-es/topics/large-language-models>.

- [9] EULA, *Acuerdo de Licencia de Usuario Final*, jul. de 2023. dirección: <https://olioapp.com/es/acuerdo-de-licencia-usuario-final/>.
- [10] C. Cadwalladr y E. Graham-Harrison, *Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach*, 17 de mar. de 2018. dirección: <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>.
- [11] Powerdata, *GDPR: Lo que debes saber sobre el reglamento general de protección de datos*, <https://www.powerdata.es/gdpr-proteccion-datos>, mayo de 2020.
- [12] S. Wang, T. Hu, H. Xiao et al., *GPT, large language models (LLMs) and generative artificial intelligence (GAI) models in geospatial science: a systematic review*, 2024. doi: 10.1080/17538947.2024.2353122.
- [13] W. X. Zhao, K. Zhou, J. Li et al., "A Survey of Large Language Models," mar. de 2023. dirección: <http://arxiv.org/abs/2303.18223>.
- [14] *Modelos De Lenguaje Grandes (LLM) | MongoDB*. dirección: <https://www.mongodb.com/es/resources/basics/artificial-intelligence/large-language-models>.
- [15] T. Ivănușcă y C. I. Irimia, "The Impact of Prompting Techniques on the Security of the LLMs and the Systems to Which They Belong," *Applied Sciences (Switzerland)*, vol. 14, 19 oct. de 2024, issn: 20763417. doi: 10.3390/app14198711.
- [16] Z. Zheng, K. Ning, Q. Zhong et al., "Towards an Understanding of Large Language Models in Software Engineering Tasks," *IEEE Computer Society*, págs. 255-257, ago. de 2023, issn: 02705257. doi: 10.1145/3639478.3639798. dirección: <http://arxiv.org/abs/2308.11396>.
- [17] Molina Peña María Elena, "Análisis automático de obligaciones y derechos en acuerdos de cliente," *Lenguas Modernas*, n.º 62, págs. 183-209, 2023, issn: 07195443. doi: 10.13039/501100011033.
- [18] *Procesamiento de lenguaje natural y sus aplicaciones (I) - Tecno Blog*. dirección: <https://www.tecnoblog.org/desarrollo/procesamiento-lenguaje-natural/>.

- [19] S. Anno, Y. Kimura y S. Sugita, "Using transformer-based models and social media posts for heat stroke detection," *Scientific Reports*, vol. 15, n.º 1, dic. de 2025, issn: 20452322. doi: 10.1038/s41598-024-84992-y.
- [20] C. D. Manning, "Human Language Understanding & Reasoning," *Daedalus*, vol. 151, n.º 2, págs. 127-138, mayo de 2022, issn: 15486192. doi: 10.1162/DAED-a-01905.
- [21] D. Myers, R. Mohawesh, V. I. Chellaboina et al., *Foundation and large language models: fundamentals, challenges, opportunities, and social impacts*, feb. de 2024. doi: 10.1007/s10586-023-04203-7.
- [22] Peña Molino Elena y García María José, "Towards a Systematic Comparison Framework for Cloud Services Customer Agreements," *Lenguas Modernas*, n.º 62, págs. 183-209, 2023, issn: 07195443. doi: 10.1007/978-981-97-0989-2_19.
- [23] E. Molino-Peña y J. M. García, "Limitaciones de las técnicas actuales para detectar obligaciones y derechos en SLA," *Lenguas Modernas*, págs. 59-66, 62 2023, issn: 07195443. doi: 10.5220/0005873100590066.
- [24] J. Varas, B. V. Coronel, I. Villagrán et al., "Innovations in surgical training: exploring the role of artificial intelligence and large language models (LLM)," *Revista do Colegio Brasileiro de Cirurgioes*, vol. 50, 2023, issn: 18094546. doi: 10.1590/0100-6991E-20233605-EN.
- [25] *LM Studio - Discover, download, and run local LLMs*. dirección: https://lmstudio.ai/?utm_source=chatgpt.com.
- [26] *LM Studio: Cómo Usar LLMs de Forma Privada y Segura en tu Escritorio - IABOX-TOOL*. dirección: https://iaboxtool.es/lm-studio-usar-llms-forma-privada-segura-escritorio/?utm_source=chatgpt.com.
- [27] *LM Studio Características, Pros, Contras y Casos de Uso*. dirección: https://aipure.ai/es/products/lm-studio/features?utm_source=chatgpt.com.

- [28] *Ejecuta modelos de IA Generativa en tu ordenador: guía paso a paso para instalar LM Studio*. dirección: https://telefonicatech.com/blog/guia-para-instalar-lm-studio?utm_source=chatgpt.com.
- [29] P. S. Moreno y D. B. Bastidas, “Impacto de la Ley Orgánica de Protección de Datos Personales en las Entidades Reguladas por la Superintendencia de la Economía Popular y Solidaria,” mayo de 2023, Consultado el 5 de febrero de 2025. dirección: <https://www.solines.ec>.
- [30] I. Ecuador, “Ley Orgánica de Protección de Datos Personales,” mayo de 2021. dirección: <https://www.importacionesecuador.com.ec/normativas/ley-organica-de-proteccion-de-datos-personales>.
- [31] I. Ozkaya, *Application of Large Language Models to Software Engineering Tasks: Opportunities, Risks, and Implications*, mayo de 2023. doi: 10.1109/MS.2023.3248401.
- [32] Y. Lu, A. Aleta, C. Du, L. Shi e Y. Moreno, *LLMs and generative agent-based models for complex systems research*, dic. de 2024. doi: 10.1016/j.plrev.2024.10.013.
- [33] “Efectos de las innovaciones en la industria Technological Innovation Project cuadernos delecenter,” inf. téc., 2007. dirección: www.ebcenter.org/proyectos.
- [34] S. Huang, Y. Huang, Y. Liu, Z. Luo y W. Lu, “Are large language models qualified reviewers in originality evaluation?” *Information Processing and Management*, vol. 62, n.º 3, mayo de 2025, issn: 03064573. doi: 10.1016/j.ipm.2024.103973.
- [35] A. Fan, B. Gokkaya, M. Harman et al., “Large Language Models for Software Engineering: Survey and Open Problems,” *IEEE Explorer*, oct. de 2023. doi: 10.1109/ICSE-FoSE59343.2023.00008. dirección: <http://arxiv.org/abs/2310.03533>.
- [36] T. Seabrooke, E. Schneiders, L. Dowthwaite et al., “A Survey of Lay People’s Willingness to Generate Legal Advice using Large Language Models (LLMs),” en *ACM International Conference Proceeding Series*, Association for Computing Machinery, sep. de 2024, isbn: 9798400709890. doi: 10.1145/3686038.3686043.

- [37] Y. Sun, C. Liu, X. Deng et al., “Development of an Intelligent Design and Simulation Aid System for Heat Treatment Processes Based on LLM,” en *28th IFHTSE 2023 Congress*, International Federation for Heat Treatment y Surface Engineering, IFHTSE, 2023, isbn: 9781713889533. doi: 10.1016/j.matdes.2024.113506.
- [38] S. Bird, E. Loper y E. Klein, *NLTK: Natural Language Toolkit*, Accessed: 2025-03-16, 2025. dirección: <https://www.nltk.org/>.
- [39] S. Ramírez, *FastAPI: High performance, easy to learn, fast to code, ready for production*, Accessed: 2025-03-16, 2025. dirección: <https://fastapi.tiangolo.com/>.
- [40] G. van Rossum y the Python development team, “Python Tutorial Release 3.13.2,” inf. téc., 2025.
- [41] A. Ravichander, A. Black, S. Wilson, T. Norton y N. Sadeh, *Question Answering for Privacy Policies: Combining Computational and Legal Perspectives*, 2019. dirección: <https://www.mturk.com/>.
- [42] J. D. Hernández, “Sobre el uso adecuado del coeficiente de correlación de Pearson: definición, propiedades y suposiciones,” vol. 37, 2018. dirección: <https://orcid.org/0000-0003->.

ANEXOS

Diagrama que muestra la estructura en capas del sistema, incluyendo la capa de presentación (frontend), capa de API (backend), capa de servicios y almacenamiento.

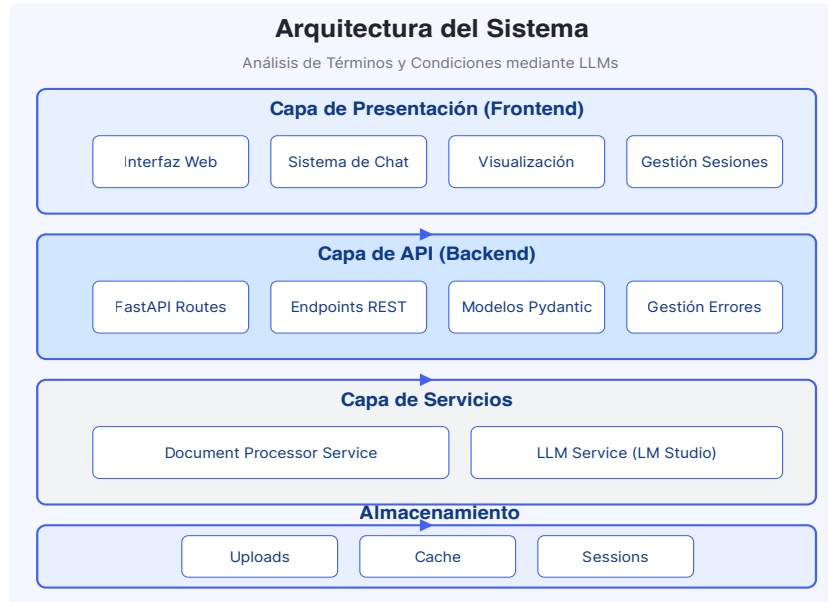


Fig. 16.

Arquitectura General del Sistema de Análisis de Términos y Condiciones

Organización jerárquica de carpetas y archivos del sistema, mostrando la distribución de componentes en el directorio principal PROYECTO_TESIS/.

Estructura de Directorios

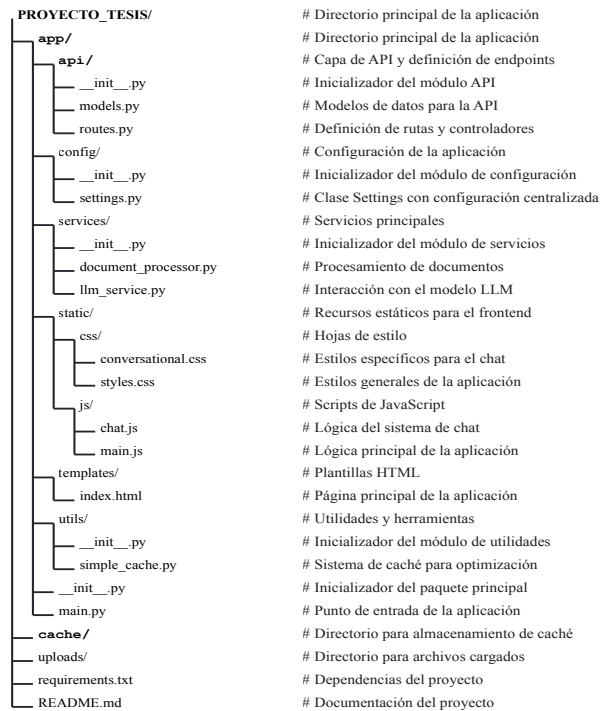


Fig. 17.

Estructura de Directorios del Proyecto

Captura de pantalla de terminal que muestra la ejecución del sistema TC Analyzer mediante el comando "python -m app.main". Se observa la inicialización exitosa del servicio LLM con el modelo hugging-quants/llama-3.2-3b-instruct, la configuración del sistema de caché (habilitado) y la asignación de 8 hilos máximos para procesamiento paralelo. La terminal también muestra la inicialización del servidor Uvicorn en <http://localhost:8000>, indicando que la aplicación está lista para recibir peticiones.

```
projecto_tesis — Python — Python -m app.main — 95x29
Last login: Sat May 3 16:55:32 on console
leoluna@MacBook-Air-de-Leoluna proyecto_tesis % /Users/leoluna/Desktop/proyecto_tesis/llms/bin/
python -m app.main
INFO:llm_service:Iniciado LLMServicio con modelo: hugging-quant/llama-3.2-3b-instruct
INFO:llm_service:Cache habilitado: True
INFO:llm_service:Threads máximos: 8
INFO: Will watch for changes in these directories: ['/Users/leoluna/Desktop/proyecto_tesis'
]
INFO: Uvicorn running on http://localhost:8000 (Press CTRL+C to quit)
INFO: Started reloader process [4054] using StatReload
INFO:llm_service:Iniciado LLMServicio con modelo: hugging-quant/llama-3.2-3b-instruct
INFO:llm_service:Cache habilitado: True
INFO:llm_service:Threads máximos: 8
INFO: Started server process [4057]
INFO: Waiting for application startup.
INFO: Application startup complete.
```

Fig. 18.
Iniciación del Sistema en Entorno Local

Captura de pantalla de la página principal del sistema, mostrando el área de carga de documentos, el panel de chat y la visualización de resultados.

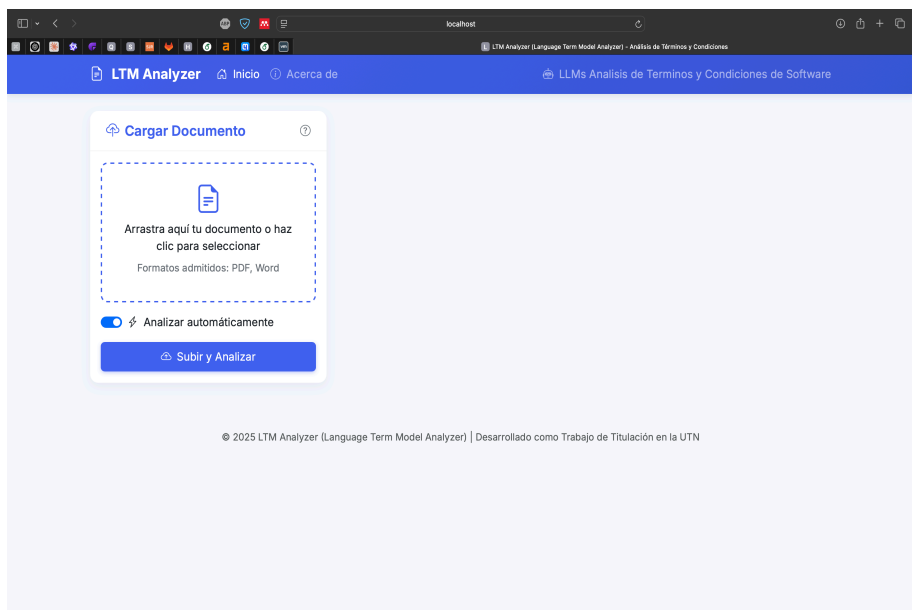


Fig. 19.
Interfaz de Usuario: Pantalla Principal

Ventana de diálogo del sistema operativo para seleccionar un documento de términos y condiciones a analizar, mostrando varios archivos PDF de diferentes proveedores de servicios tecnológicos como GitLab, Google Workspace y Microsoft Teams.

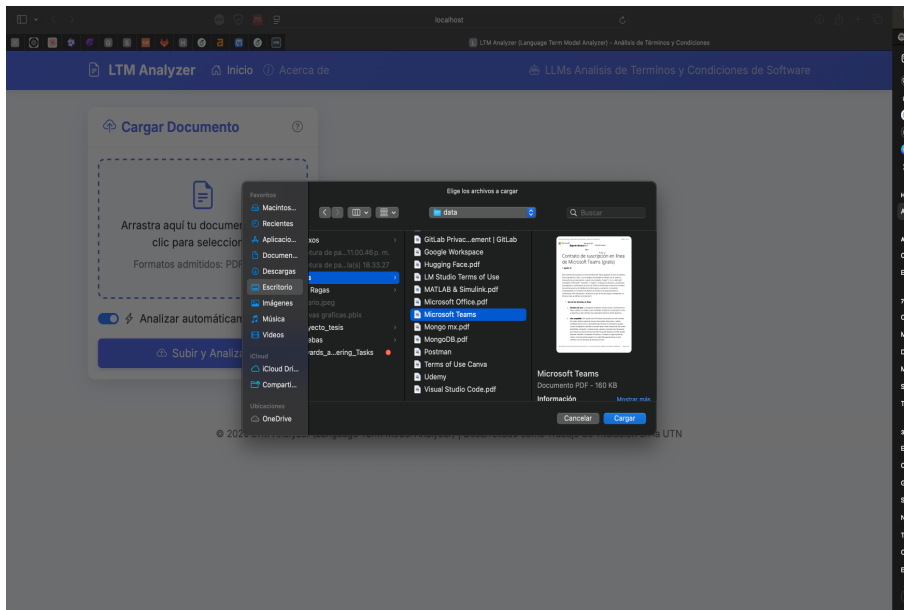


Fig. 20.
Selección de Documentos para Análisis

Ventana modal mostrando el proceso de carga del archivo "Postman.pdf" con un indicador de progreso circular y barra de estado. El sistema informa que el documento está siendo procesado mientras se visualiza en segundo plano la interfaz principal con el panel de chat y la confirmación de carga exitosa.

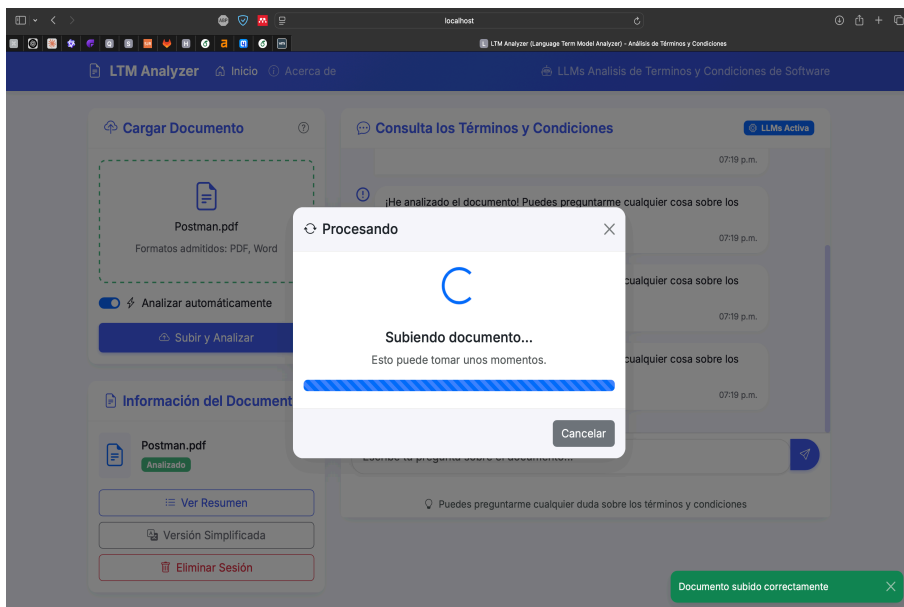


Fig. 21.
Interfaz de Carga de Documentos del Sistema

Interfaz de LTM Analyzer con un documento de Postman procesado, donde aparece desplegada la ventana Resumen del Documento”. El panel izquierdo contiene opciones para consultar el resumen, crear una versión simplificada o eliminar la sesión activa. El panel derecho exhibe el resumen elaborado con estructura organizada de los elementos centrales del documento: requisitos de aceptación, restricciones etarias y seguridad de credenciales.

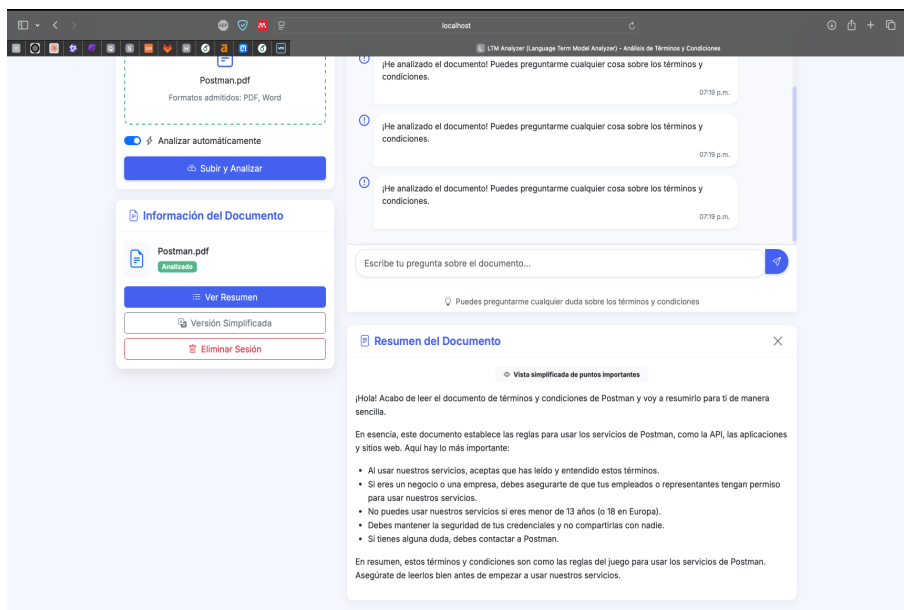


Fig. 22.

Visualización del Resumen de Términos y Condiciones

Interfaz con la versión simplificada de términos y condiciones de Postman que LTM Analyzer generó automáticamente. El contenido aparece organizado estructuralmente usando lenguaje accesible, donde resaltan cinco elementos centrales: requisitos etarios, condiciones de uso, gestión de cuentas, trabajo colaborativo y restricciones de API. Además incorpora apartados sobre modificaciones a los términos y por qué resulta importante revisarlos previo al uso del servicio.

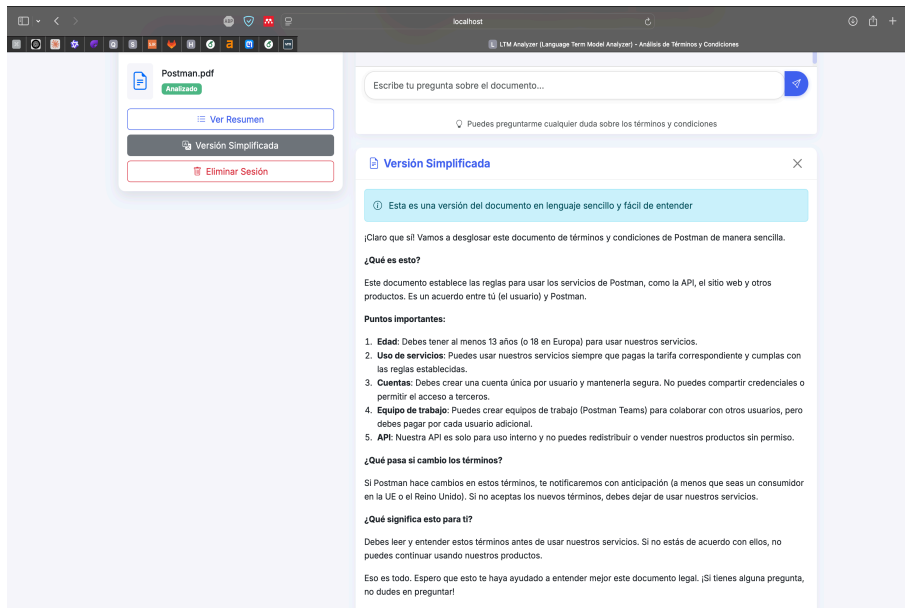


Fig. 23.
Versión Simplificada de Términos y Condiciones

Interfaz de chat de LTM Analyzer mostrando una consulta específica sobre los términos de licencia de Postman y la respuesta generada por el modelo. El usuario ha preguntado si ”¿La licencia otorgada al usuario final es de tipo perpetua, temporal o revocable, y bajo qué condiciones puede rescindirse unilateralmente?”, y el sistema ha proporcionado una respuesta detallada explicando que la licencia es revocable y las condiciones bajo las cuales el usuario puede rescindir el acuerdo. La interfaz muestra la aplicación funcionando correctamente con el indicador ”LLMs Activa.^{en} la parte superior.

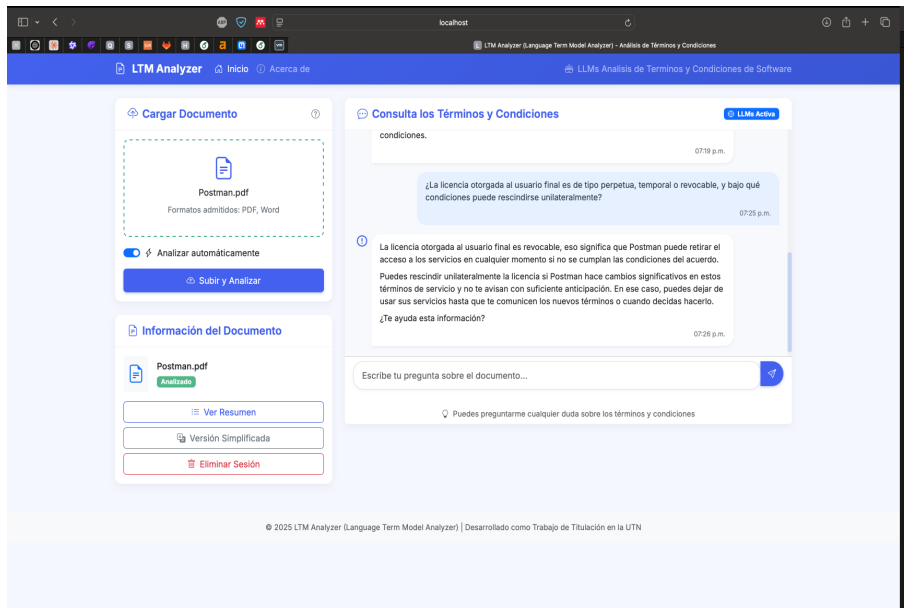


Fig. 24.
Conversación sobre Términos de Licencia

Captura de pantalla del entorno LM Studio mostrando el modelo hugging-quant/llama-3.2-3b-instruct en ejecución, mostrando los resultados que se hacen en la interfaz web.

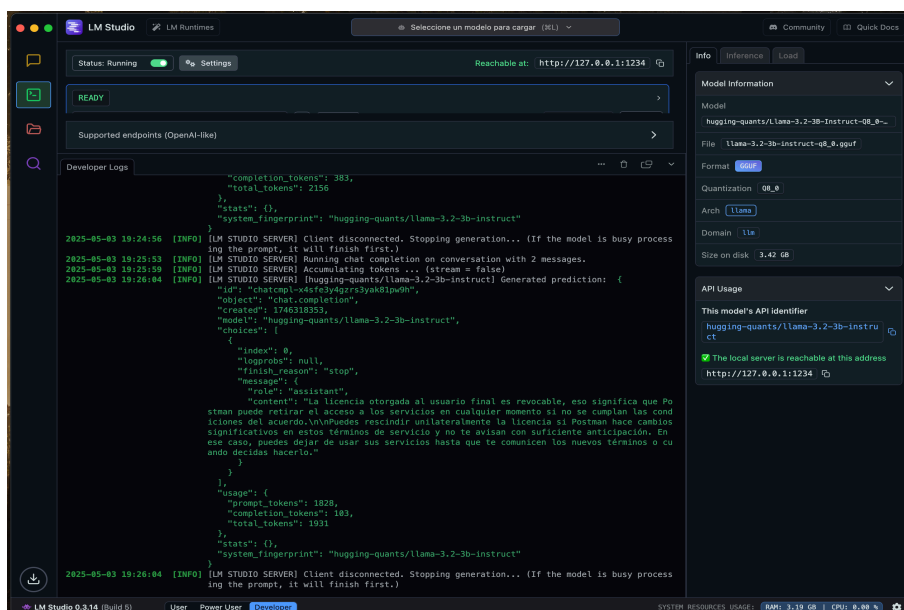


Fig. 25.
Interfaz de LM Studio con Modelo Llama-3.2-3b-instruct

Interfaz de LTM Analyzer mostrando una consulta sobre restricciones de ingeniería inversa en Postman. El sistema responde explicando que esta práctica está generalmente prohibida sin

consentimiento explícito, aunque existe una excepción para software de código abierto que se rige por sus propias licencias.

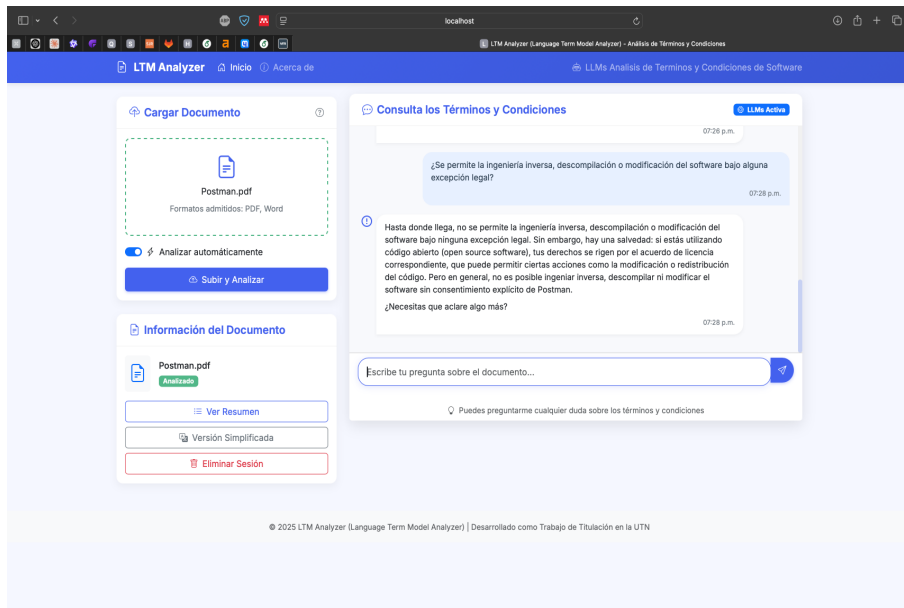


Fig. 26.
Consulta sobre Ingeniería Inversa en Términos Legales

Directorio de caché con múltiples archivos JSON identificados por hashes y organizados cronológicamente. Estos archivos almacenan resultados de análisis previos para optimizar la velocidad de respuesta en consultas repetidas, mejorando así el rendimiento general del sistema.

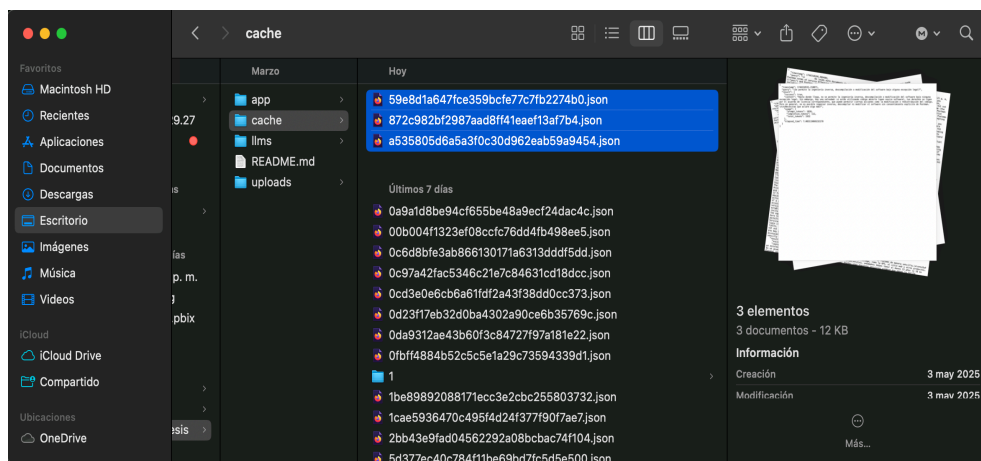


Fig. 27.
Sistema de Caché del Analizador de Términos y Condiciones

Contenido muestra el contenido de un archivo JSON almacenado en la caché del sistema LTM Analyzer con información detallada.

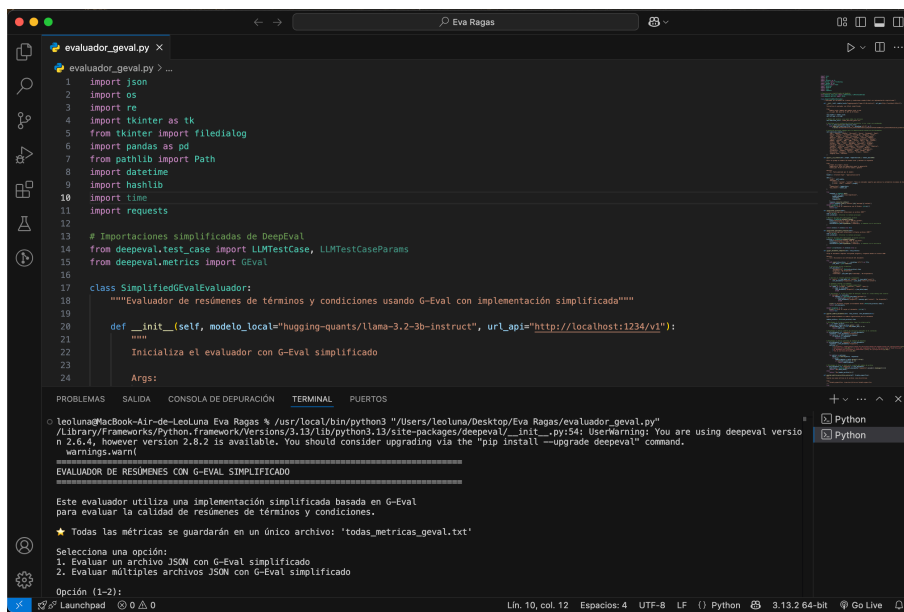


A screenshot of a browser window showing a JSON file named `872c982bf2987aad8ff41eaf13af7b4.json`. The JSON content is as follows:

```
{
  "timestamp": 1746318364.7681311,
  "query": "¿La licencia otorgada al usuario final es de tipo perpetua, temporal o revocable, y bajo qué condiciones puede rescindirse unilateralmente?",
  "result": {
    "success": true,
    "content": "La licencia otorgada al usuario final es revocable, eso significa que Postman puede retirar el acceso a los servicios en cualquier momento si no se cumplen las condiciones del acuerdo.\n\nPuedes rescindir unilateralmente la licencia si Postman hace cambios significativos en estos términos de servicio y no te avisan con suficiente anticipación. En ese caso, puedes dejar de usar sus servicios hasta que te comuniquen los nuevos términos o cuando decidas hacerlo.\n\n¿Te ayuda esta información?",
    "usage": {
      "prompt_tokens": 1028,
      "completion_tokens": 103,
      "total_tokens": 1931
    },
    "elapsed_time": 11.4494469165802
  }
}
```

Fig. 28.
Archivo de Caché JSON del Sistema

Código en Python que implementa una versión simplificada de GEval para evaluar resúmenes de documentos legales generados por LTM Analyzer. El script utiliza el modelo llama-3.2-3b-instruct y ofrece opciones para evaluar archivos JSON individuales o en lote.



A screenshot of a VS Code editor window showing a Python script named `evaluador_geval.py`. The script imports various libraries and defines a `SimplifiedGEvalEvaluador` class. The terminal output shows the following:

```
PROBLEMAS SALIDA CONSOLA DE DEPURACIÓN TERMINAL PUERTOS
o leoluna@MacBook-Air-de-Leoluna: Eva Ragas % /usr/local/bin/python3 /Users/leoluna/Desktop/Eva Ragas/evaluador_geval.py
/Library/Frameworks/Python.framework/Versions/2.13/lib/python3.13/site-packages/deepeval/_init_.py:54: UserWarning: You are using deepeval version 2.8.4, however version 2.8.2 is available. You should consider upgrading via the "pip install --upgrade deepeval" command.
warnings.warn()

=====
EVALUADOR DE RESÚMENES CON G-EVAL SIMPLIFICADO
=====

Este evaluador utiliza una implementación simplificada basada en G-Eval para evaluar la calidad de resúmenes de términos y condiciones.

★ Todas las métricas se guardarán en un único archivo: 'todas_mtricas_geval.txt'

Selecciona una opción:
1. Evaluar un archivo JSON con G-Eval simplificado
2. Evaluar múltiples archivos JSON con G-Eval simplificado

Opción (1-2):
```

Fig. 29.
Script Evaluador de Análisis de Términos y Condiciones

Interfaz de VS Code con el evaluador GEval ejecutándose, donde aparece un diálogo de selección de archivos JSON desde la caché. Estos archivos contienen respuestas procesadas sobre licencias de Postman destinadas a evaluación sistemática.

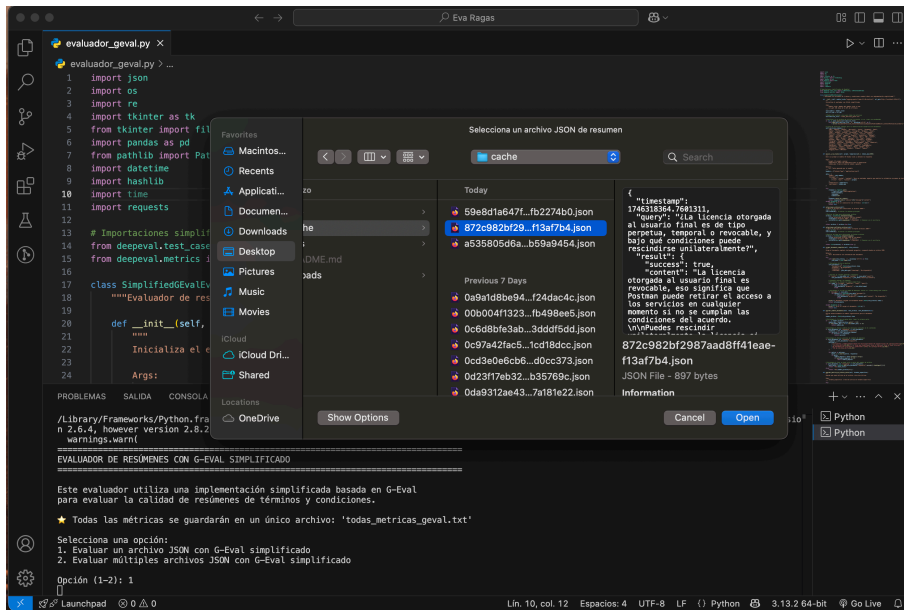


Fig. 30.
Selección de Archivo de Caché en Evaluador GEval

Terminal de VS Code con resultados de evaluación GEval aplicada al análisis de términos y condiciones. La evaluación presenta métricas detalladas correspondientes al documento "Doc5752e0z" almacena los resultados en tres archivos distintos para revisión posterior.

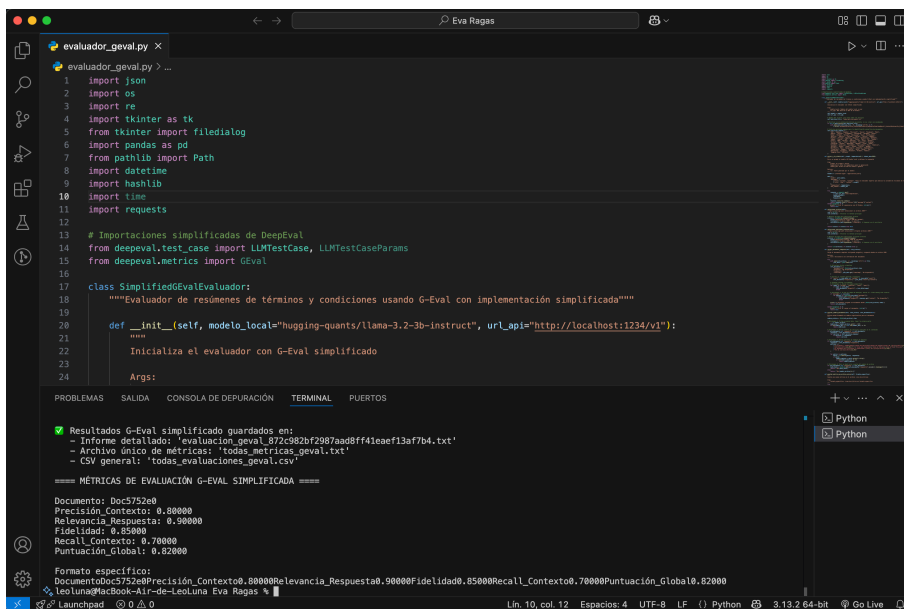


Fig. 31.
Resultados de Evaluación GEval para Documentos Legales

Archivo de texto con resultados detallados de evaluación GEval aplicada al análisis de términos de licencia de Postman. Incluye métricas de calidad junto con justificación completa que identifica fortalezas y limitaciones del resumen elaborado.

```

evaluacion_geval_872c982bf2987aad8ff41eae13af7b4.txt
Abrir con Editor de Texto

EVALUACIÓN SIMPLIFICADA G-EVAL DEL RESUMEN DE TÉRMINOS Y CONDICIONES
=====
DOCUMENTO: Doc5752e0
ID DEL DOCUMENTO: 872c982bf2987aad8ff41eae13af7b4
FECHA DE EVALUACIÓN: 2025-05-03 19:36:44

PREGUNTA ORIGINAL:
¿La licencia otorgada al usuario final es de tipo perpetua, temporal o revocable, y bajo qué condiciones puede rescindirse unilateralmente?

RESUMEN EVALUADO:
La licencia otorgada al usuario final es revocable, eso significa que Postman puede retirar el acceso a los servicios en cualquier momento si no se cumplen las condiciones del acuerdo.

Puedes rescindir unilateralmente la licencia si Postman hace cambios significativos en estos términos de servicio y no te avisan con suficiente anticipación. En ese caso, puedes dejar de usar sus servicios hasta que te comuniquen los nuevos términos o cuando decidas hacerlo.

¿Te ayuda esta información?

MÉTRICAS DE EVALUACIÓN G-EVAL SIMPLIFICADA:
-----
Precisión_Contexto: 0.80000
Relevancia_Respuesta: 0.90000
Fidelidad: 0.85000
Recall_Contexto: 0.70000
Puntuación_Global: 0.82000

JUSTIFICACIÓN DE LA EVALUACIÓN G-EVAL:
-----
En general, el resumen muestra una buena precisión al capturar la información clave del documento original. Sin embargo, hay algunas limitaciones en términos de relevancia y recall. La pregunta planteada se centra en la posibilidad de ingeniería inversa, descompilación o modificación del software, pero el resumen no aborda todas las posibles excepciones o situaciones en las que se podría permitir estas acciones. Aunque la frase "sin consentimiento explícito de Postman" puede ser un problema, ya que podría interpretarse como una restricción adicional no mencionada en el documento original. Además, el resumen no captura todos los puntos importantes del documento original, como la política de privacidad o la responsabilidad de Postman en caso de violaciones de derechos de autor.

Aunque en términos de relevancia, el resumen responde específicamente a la pregunta planteada, pero no aborda todas las posibles situaciones en las que se podría permitir la ingeniería inversa, descompilación o modificación del software. En este sentido, el resumen es un buen punto de partida para entender las restricciones establecidas por Postman, pero requiere más información para ser considerado completo y preciso.

Aunque en términos de fidelidad y precisión, el resumen muestra una buena precisión y fidelidad al contenido del documento original, pero tiene algunas limitaciones en términos de relevancia y recall. Requiere más información para ser considerado completo y preciso, y puede requerir ajustes para abordar todas las posibles excepciones o situaciones en las que se podría permitir la ingeniería inversa, descompilación o modificación del software.

```

Fig. 32. Informe Detallado de Evaluación GEval para Análisis Legal

Hoja de cálculo con resultados consolidados de evaluaciones GEval múltiples sobre documentos de términos y condiciones de varios proveedores de servicios. Cada fila corresponde a una consulta específica, mientras las columnas exhiben las métricas de calidad respectivas.

DocId	DocTitle	DocContent	Pregunta	Resumen	Precisión	Relevancia	Fidelidad	Recall	Puntuación
DocId4821	0a76f1a50e75bac3f84a3c5336632f1	¿En qué circunstancias puede el proveedor suspender o terminar el servicio?	0.8	0.9	0.7	0.6	0.75		
DocId8c281	8a29a9701c06eb85481e632021b23ad	¿qué beneficios tengo como educador	0.8	0.9	0.7	0.6	0.75		
GitLab	0b617478b10715e720c630448798a4	¿Hay alguna limitación de responsabilidad mencionada en el documento? ¿Cuáles son?	0.9	0.8	0.85	0.7	0.82		
GitLab	d17829c7c763eb3369a1986292ef	¿Cuáles son las políticas de reembolso según estos términos?	0.8	0.9	0.6	0.4	0.7		
DocIdf004	62760599840c202206493102c394e	¿Qué dice el documento sobre resolución de disputas o arbitraje?	0.0	0.0	0.0	0.0	0.0		
Enstr	12a85f589a4321014d32a0a59070cc	¿Existe alguna cláusula sobre modificaciones futuras a los términos y condiciones?	0.9	1.0	0.8	0.7	0.85		
Google	3bac57423206edc70031c8844250f	¿Qué restricciones de uso se mencionan en el documento?	0.0	0.0	0.0	0.0	0.0		
Google	99920337684d0f0082575f2430a0f8	¿Cuáles son las obligaciones del proveedor hacia el usuario según estos términos?	0.9	1.0	0.8	0.7	0.85		
Google	8d9eab7139e8df540a3a3e4b16504	¿Hay alguna mención sobre la ley aplicable o jurisdicción en caso de conflictos?	0.9	0.8	0.85	0.7	0.82		
Google	0a473dc0530b47e40485a24295f8	¿Existen cláusulas relacionadas con la seguridad de la información?	0.9	1.0	0.0	0.0	0.0		
Meta	880c98350c753346489990339d00f	¿Cuáles son las condiciones para la transferencia de datos a terceros?	0.8	0.9	0.7	0.6	0.75		
Meta	b7868d23f70c806bea95e31912107	¿Qué dice el documento sobre la duración del contrato?	0.8	0.9	0.6	0.4	0.65		
Meta	3a8248f9f6e20c020c0b0a018a55	¿Existen requisitos de edad mínima para utilizar el servicio?	0.0	0.0	0.0	0.0	0.0		
Adobe	b7640970e48a9a7a75c74c3080a0	¿Existen requisitos de edad mínima para utilizar el servicio?	0.9	0.8	0.85	0.7	0.82		
Adobe	758c28a996d5c3f0d1879c189f113	¿Hay alguna cláusula abusiva o potencialmente problemática en estos términos y condiciones?	0.0	0.0	0.0	0.0	0.0		
Adobe	04af839300f5c84a0f9e3d5c73	Resume los principales puntos de estos términos y condiciones en lenguaje simple.	0.8	0.0	0.0	0.0	0.0		
Hugging Face	52904223998571956894a48306e10	Resume los principales puntos de estos términos y condiciones en lenguaje simple.	0.8	1.0	0.8	0.7	0.85		
Mathworks	41130f3f9531ae5a09c784e99004	¿qué venecios y desventajas hay	0.9	0.9	0.7	0.6	0.75		
Mathworks	ea2865a0f0b0aa5973a87b74c0a0af	¿darne un resumen de los puntos mas importantes	0.8	0.9	0.7	0.6	0.75		
Hugging Face	87106344f8f28162607499683c5cd0	¿Existen cláusulas sobre actualizaciones automáticas del software?	0.8	0.9	0.7	0.6	0.75		
Zoom	073110a70732a0f0974971318a612	¿Cuáles son las políticas de privacidad en caso de una violación de datos?	0.8	0.9	0.7	0.6	0.75		
Zoom	660f0e2904b0a37032a4ead48f68	¿Qué menciona el documento sobre la portabilidad de datos? ¿Puede el usuario exportar su información...	0.8	0.9	0.6	0.4	0.75		
Zoom	66d059887c74727281f5188a50700	¿Existen cláusulas relacionadas con la supervisión o monitoreo del uso del software?	0.8	0.9	0.8	0.7	0.85		
Zoom	495060daa4270668076032c2988494	¿Dame mas información sobre la sección 15	0.8	0.9	0.7	0.6	0.75		
MongoDBMX	179c19fc3248290191199bb272a4c	de que se trata el documento	0.9	0.0	0.0	0.0	0.0		
MongoDBMX	179c19fc3248290191199bb272a4c	de que se trata el documento	0.8	0.9	0.85	0.0	0.0		
MongoDBMX	1e51407852bd451441e0d5b5a617a	¿hay alguna violación de datos	0.8	0.9	0.7	0.0	0.0		
Postman	b5662a6f782c3a0202ae40212f305	¿Qué menciona el documento sobre el uso del software en diferentes dispositivos o instalaciones?	0.8	0.9	0.7	0.6	0.75		
Postman	4878781116909297490a08080734	¿Cuáles son las políticas de privacidad en caso de una violación de datos?	0.8	0.9	0.7	0.6	0.75		
Postman	6953f56253116999028370140a0a08	¿me puedes dar mas información de la sección 3.2 sobre el uso de software en diferentes dispositivos	0.8	0.8	0.85	0.7	0.82		
Microsoft	a8829433ac88b150ca3666d6ca810	¿Qué menciona el documento sobre la compatibilidad con software de terceros?	0.0	0.0	0.0	0.0	0.0		
Microsoft	46c400110260786c78a9f9322860e	¿dame mas información sobre eso	0.8	0.9	0.7	0.6	0.75		
LMStudio	0dca3d789908a052690f04021033a	¿Cuáles son las políticas relacionadas con cuentas inactivas o abandonadas?	0.8	1.0	0.8	0.7	0.85		
LMStudio	0f349ff97748f63181323467634	¿El documento incluye información sobre cómo el usuario puede acceder al soporte técnico?	0.8	0.9	0.7	0.6	0.75		
LMStudio	6e8828d6bc3f809e66c0bc0e27f7	¿Me puedes dar mas información sobre como el usuario puede acceder al soporte técnico	0.9	0.8	0.85	0.7	0.82		
LMStudio	09f17e2204453a7a61c217138c38	¿Hay alguna cláusula sobre la revocación del consentimiento por parte del usuario?	0.9	1.0	0.8	0.7	0.85		
LMStudio	c8b6d880c38488a3a6c07289756	¿Si dime lo referente a la cláusula	0.9	0.8	0.85	0.7	0.82		
ANS	73564679306379357ac207882d49d	¿me puedes decir qué beneficios tengo de este documento como cliente	0.8	0.9	0.6	0.4	0.7		
DocId4821	0a76f1a50e75bac3f84a3c5336632f1	¿En qué circunstancias puede el proveedor suspender o terminar el servicio?	0.8	0.9	0.7	0.6	0.75		
resultados	resultados_evaluacion	No disponible	0.8	0.6	0.9	0.7	0.75		
Doc5752e0	872c982bf2987aad8ff41eae13af7b4	¿La licencia otorgada al usuario final es de tipo perpetua, temporal o revocable, y bajo qué condi...	0.8	0.9	0.85	0.7	0.82		

Fig. 33. Archivo CSV de Evaluaciones GEval para Análisis de Términos y Condiciones

Modal del Sistema de Análisis Automático de Términos y Condiciones donde se describen sus cuatro funciones centrales: análisis automático de documentos legales, resumen de cláusulas relevantes, interfaz conversacional para consultas y simplificación del lenguaje legal hacia términos accesibles.

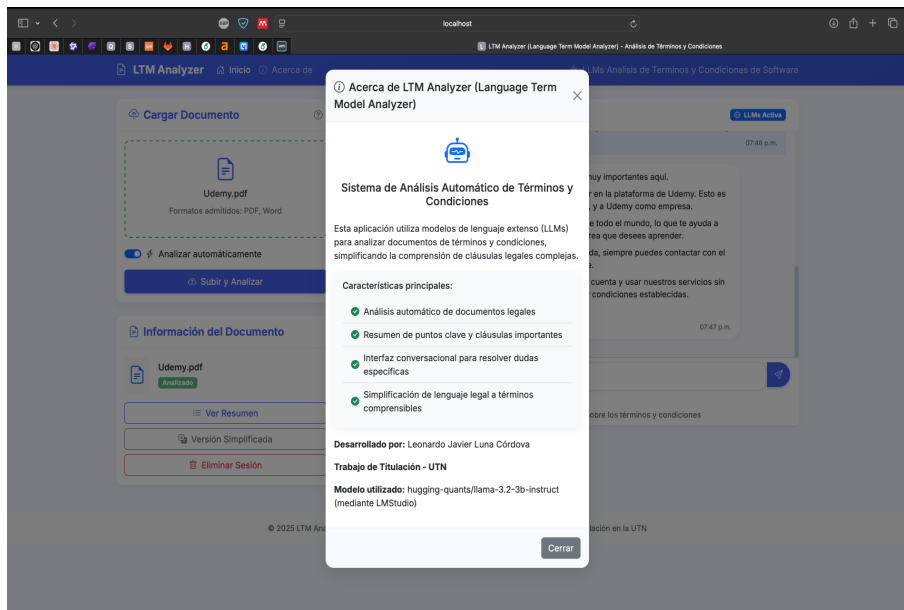


Fig. 34.
Ventana de Información del Sistema de Análisis