LLM Analysis of Legal Documents

Sankalp Shubham
sankalp.shubham@utdallas.edu
University of Texas at Dallas
Richardson, Texas, USA

Aarian Ahsan
aarian.ahsan@utdallas.edu
University of Texas at Dallas
Richardson, Texas, USA

ABSTRACT

Legal documents are often complex and dense, making them hard for non-experts to understand. This study explores how AI advancements, particularly with models like Llama-2 and GPT-4, can clarify legal texts through summarization and QA methods. We assess these models' ability to handle legal terms, reduce errors, and aid layperson understanding while considering the ethics of AI-generated legal advice. Our research highlights LLMs' potential to improve legal information access, stressing the need for accuracy, proper training, and ethical practices in AI's legal applications.

Sankalp Shubham and Aarian Ahsan. 2024. LLM Analysis of Legal Documents.

1 INTRODUCTION

The intricate nature, complex language, and specialized terms found in legal documents create considerable obstacles for those lacking expertise in this domain. These challenges can make individuals uneasy when encountering contracts, agreements, or compliance papers, often resulting in confusion, misinterpretation, or the need for costly legal consultation. The digitization of legal operations and the growing volume of legal documents worsen this issue. In 2020, the legal services market was valued around $1,000 billion, highlighting the urgent need for effective solutions to this challenge (cite source). Several tools have been developed to simplify the comprehension of legal papers, including contract analyzers and rule-based legal bots, yet they frequently fall short in offering clear and easily understandable interpretations. These solutions typically depend on fixed templates or specialized legal jargon, which limits their adaptability to a wider range of legal texts. Additionally, they often lack the ability to engage in interactive dialogues with users, reducing their efficiency in resolving individual questions or accommodating unique document types.

Regarding nonstandard documents or complex clauses, these tools often provide only limited insights, leaving users with unanswered questions. Given the constraints of existing solutions, there is a growing interest in utilizing artificial intelligence, especially large language models (LLMs) capable of comprehending and generating natural language. Models like GPT-4 and LLaMA-2 offer significant promise in making complex texts more understandable across various fields. Nonetheless, their application in the legal field is still evolving, paving the way for innovative progress. Preliminary research has shown encouraging results in tasks such as summarizing, emphasizing key aspects, and responding to queries, suggesting that LLMs may improve the understanding of legal documents. **Problem to be Solved**: This paper addresses the challenges non-experts encounter when interpreting legal documents. These texts often contain technical and jargon that can be difficult to navigate without legal expertise. We aim to develop a system that uses legal data-trained LLMs to provide simplified summaries, interactive question-answering, and comprehensive document analysis, empowering nonexperts to engage more confidently with legal documents.

, ,

1.1 Proposed Solution

We propose a specialized tool for legal document analysis and interaction through a conversational interface. Users can upload documents to receive targeted insights, including summaries, highlighted sections, and answers to specific queries. Unlike general-purpose models like ChatGPT, our solution is fine-tuned for legal contexts, ensuring accuracy, reliability, and legal compliance. Key components include:

Document Upload and Parsing: Users upload legal documents, such as contracts or regulatory filings, for domain-specific analysis. Parsing identifies legal terminology and structures, optimizing the model to focus on relevant elements.

Legal Document Summarization: Our tool summarizes complex legal texts using fine-tuned LLMs, emphasizing critical clauses, obligations, and rights. This ensures legally contextualized summaries that simplify language while preserving essential details and accuracy.

Interactive Question-Answering (Q&A): The Q&A functionality answers legal queries with contextually accurate and legally grounded responses, informed by case law and domain-specific data. This offers a deeper, law-informed perspective compared to general-purpose models like ChatGPT.

Error Detection and Source Validation: To address hallucinations, our tool includes error detection and real-time validation. Feedback mechanisms with confidence scores enhance reliability and reduce risks of misleading outputs, setting it apart from general models.

Customization for Legal Domains: Fine-tuned on legal corpora, our solution provides insights across domains like contract law, compliance, and intellectual property. For instance, it highlights critical sections such as indemnity or confidentiality clauses, offering precise analysis beyond the capability of generic LLMs.

By focusing exclusively on legal applications, our tool delivers precise and compliant responses, offering a dependable alternative to general-purpose models for legal document analysis.

1.2 Contributions of the Paper

This study provides significant contributions to the field of legal document analysis. First, it presents the inaugural implementation of a fine-tuned large language model (LLM)-based tool engineered specifically for the summarization and interpretation of legal documents, created with non-expert users in mind. Second, it introduces a novel integration of question-answering functionalities within the legal document analysis process, enabling users to engage with complex texts interactively. Third, the research addresses the issue of LLM hallucinations by incorporating real-time feedback mechanisms and confidence scores, thus improving the reliability of AI output in legal contexts. Finally, an ethical framework is proposed to ensure the responsible deployment of AI, guiding users through legal content while avoiding the provision of legal advice.

2 BACKGROUND

Applying large language models (LLMs) to legal document analysis requires addressing the complexity of legal language, leveraging NLP advancements, and overcoming domain-specific challenges. This section outlines these foundational aspects to contextualize our proposed solution.

2.1 Legal Language Complexity

Legal documents use precise, formal language filled with archaic terms, technical jargon, and hierarchical structures. Rights, obligations, and terms are conveyed with strict attention to detail, where even minor wording changes can alter interpretations. Cross-references and layered clauses increase cognitive load, making legal texts challenging to interpret for non-experts.

2.2 Natural Language Processing (NLP) in Legal Contexts

NLP technologies, including LLMs like GPT-4 and LLaMA-2, have advanced tasks such as summarization and question answering. However, legal texts demand precision, with tokenization preserving the integrity of terminology. Domain-specific tuning is essential to ensure LLMs handle legal nuances effectively, overcoming generic limitations.

2.3 Challenges in Legal Document Processing

Applying LLMs to legal documents involves several challenges:
Ambiguity and Vagueness: Legal language are often intentionally vague, complicating LLM responses.
Document Length: Legal texts are typically lengthy, exceeding input token limits and requiring methods like text chunking while preserving meaning.
Legal Jargon: The use of specialized terminology necessitates exposure to legal corpora to avoid misinterpretation.
Hallucinations: LLMs risk generating plausible but incorrect information, which can have serious legal implications.

2.4 Evolution Legal Document Analysis

Traditional AI approaches to legal analysis relied on rule-based systems and basic NLP techniques like keyword extraction and text classification. While effective for specific tasks, these systems lacked flexibility and struggled with unstructured documents or complex relationships. Template-based methods worked well for structured texts but offered limited adaptability and interactivity.

The advent of LLMs, such as BERT, GPT-4, and LLaMA-2, marked a transformative shift in NLP. These models use deep neural networks to contextualize language, excelling at complex tasks like summarization and question answering. Fine-tuning on legal datasets enhances their ability to interpret intricate legal texts, surpassing traditional methods and enabling more robust document analysis.

3 RELATED WORK

The application of large language models (LLMs) to legal document analysis has gained significant attention due to the need to simplify legal terminology and enhance accessibility for non-experts. This project aims to develop a tool leveraging LLMs trained on legal corpora to provide users with document summaries and analyses, addressing challenges in complexity, legal interpretation, and comprehension. This section reviews related studies on LLM applications, hallucination mitigation, and ethical considerations in legal AI.

3.1 LLMs for Legal Document Summarization and Analysis

LLMs like LLAMA-2 and DistilBERT have demonstrated their ability to simplify complex legal texts into manageable, insightful summaries. By employing techniques such as text chunking and tokenization, these models process dense legal documents across various domains, including financial and regulatory sectors, while maintaining semantic coherence [4]. Performance metrics, such as ROUGE and BERT scores, affirm the effectiveness of domain-specific fine-tuning in improving both accuracy and user accessibility.

This aligns with the objectives of the proposed tool, which emphasizes clear, concise legal document analysis tailored for non-experts. Fine-tuning these models on legal datasets further enhances their ability to identify critical sections, such as compliance clauses and key contractual obligations, ensuring precise and user-friendly interactions.

3.2 Addressing Hallucinations and Improving Accuracy

Hallucinations, where LLMs generate misleading or incorrect information, pose a significant challenge in legal applications due to potential consequences for users. Studies propose strategies such as real-time feedback and confidence scores to help users evaluate the reliability of outputs [5]. Additionally, error detection and source verification mechanisms can further mitigate risks, ensuring transparency and trust in the system.

Incorporating multiple agents to handle queries contextually has also proven effective. Assigning specific agents to different legal domains enhances response accuracy and reduces error rates, significantly improving user satisfaction [2]. Fine-tuning models on legal corpora remains crucial to minimizing hallucinations and enhancing the reliability of responses.

3.3 Legal Compliance and Ethical Considerations

The use of AI in legal contexts raises ethical concerns, particularly regarding unauthorized legal advice. LLM tools must comply with guidelines, providing informational assistance without delivering definitive legal judgments to avoid liability risks. Ethical frameworks emphasize the importance of helping users understand legal documents and formulate questions rather than substituting professional advice [1, 5].

Furthermore, studies advocate for careful evaluation of LLM responses, considering user expertise and the societal impact of AI in legal decision-making. This ensures that tools like the proposed system remain ethically sound while providing meaningful insights to non-specialist users [1].

4 PROBLEM DEFINITION

This research addresses the challenges of applying large language models (LLMs) in legal contexts, focusing on secure, reliable, and legally compliant solutions.

4.1 Challenges with General-Purpose LLMs in Legal Contexts

Lack of Legal-Focused Training: General-purpose LLMs like ChatGPT lack domain-specific expertise, often failing to provide nuanced answers rooted in legal doctrines, case precedents, and jurisdictional regulations.

Inability to Provide Explicit Legal Advice: Legal advice demands specialized expertise, and general LLMs cannot fulfill this requirement. Models tailored to deliver legally sound responses are needed to complement professional judgment.

Hallucination and Misinterpretation Risks: LLMs may generate inaccurate or misleading information, particularly in specialized domains like law. Mitigating these risks involves grounding responses in case law and legal precedents to enhance reliability.

4.2 Data Privacy and Security Concerns

Legal data requires strict adherence to privacy regulations (e.g., GDPR, HIPAA). Using generic, cloud-based LLMs can compromise data security, necessitating an internal, secure LLM framework. Additionally, external server dependency increases the risk of unauthorized access to sensitive legal data. Deploying LLMs in a private, controlled infrastructure ensures data confidentiality and regulatory compliance.

4.3 Formalized Problem Definition

Our approach seeks to formalize and address the following core problems:

Problem 1: Insufficient Legal-Specific Knowledge in LLMs General LLMs lack a structured, law-focused framework. The solution involves training on domain-specific datasets enriched with statutes, case law, and jurisdictional information.

Problem 2: Security and Compliance in Data Handling Legal applications require secure and compliant data processing systems. The proposed internal LLM addresses this by ensuring confidentiality and adherence to privacy regulations.

4.4 Technical Challenges

The analysis of legal documents utilizing large language models (LLMs) presents a multitude of challenges: comprehending intricate legal terminology and regulations demands specialized training within the domain; minimizing errors necessitates rigorous error detection protocols and verification against authoritative sources; safeguarding privacy requires the implementation of stringent data protection measures; and achieving prompt responsiveness must be balanced with the management of computational demands. Surmounting these challenges is essential for the efficacious analysis of legal documents.

Regulatory and Ethical Considerations: The model must avoid unauthorized legal advising while adhering to ethical principles like transparency and accountability. Evaluating performance remains challenging, as traditional metrics may not fully capture legal reasoning. Overcoming these challenges is essential for creating a practical tool for legal professionals.

5 EXAMPLE

To clarify our approach, we will include examples demonstrating how our LLM processes legal documents, simplifies terminology, and extracts key information, making technical aspects easier to understand and apply.

5.1 LegalMind System and the LLM-based Legal Judgment Query System

The paper titled "LegalMind: System and the LLM-based Legal Judgment Query System" outlines an elaborate system that employs large language models (LLMs) to facilitate answering legal inquiries and forecasting legal decisions. This model, specifically trained on legal datasets, enhances the precision of judgment forecasts by retrieving pertinent legal cases and suggesting precedents. A notable innovation of this system is its capacity to handle various and intricate legal terminology, enabling non-experts to more easily understand complex legal documents. The authors highlight the model's effectiveness in several legal tasks, such as query analysis and the summarization of extensive legal texts. This system demonstrates how domain-specific LLMs can boost the effectiveness and accuracy of legal research by diminishing the cognitive demands of dealing with dense legal material [6].

5.2 LawLLM and Legal Judgment Prediction

The paper "LawLLM: Legal Judgment Prediction Using Large Language Models" examines using LLMs for legal judgment prediction. LawLLM, tailored for legal datasets, excels in retrieving cases, suggesting precedents, and summarizing texts. It achieves high accuracy in predicting outcomes and recommending precedents from past cases. The paper addresses the difficulty of interpreting complex legal language and shows how LawLLM provides clear summaries and insights, making legal jargon more understandable for non-experts, thereby enhancing legal document analysis while preserving key information. [3].

6 APPROACH

The system employs tools and frameworks to streamline the analysis and query of legal documents. The workflow includes components to help extract text from legal PDFs and integrate OpenAI GPT to provide summaries and detailed answers to legal queries.

Data Ingestion: Uploaded PDF documents are processed to extract text while preserving their structure. The extracted content undergoes preprocessing to handle formatting inconsistencies, such as line breaks and redundant whitespace, ensuring the text is clean and suitable for analysis.

Building a Knowledge Base: The system converts preprocessed text chunks into vector embeddings using OpenAI Embeddings. These embeddings are stored in a FAISS-based vector database, enabling efficient similarity-based retrieval of relevant sections during query processing. To handle lengthy legal documents, the text is divided into manageable chunks. A map-reduce summarization chain generates concise summaries for each section and combines them into a cohesive overall summary.

Document Validation: A document's authenticity and structure are validated through keyword matching, structural integrity checks, and classification using a Legal-BERT model. This step ensures the input document aligns with predefined legal standards, adding an extra layer of reliability to the analysis. If the document is valid, it is added to the existing knowledge base.

Legal Question Answering: For user queries, the system retrieves relevant text chunks from the knowledge base that has been enhanced with Retrieval Augment Generation (RAG) using similarity search. The retrieved chunks are fed into LangChain, which uses a custom-designed legal QA prompt to generate detailed, contextually accurate responses tailored to the user's query.

User Interaction: The interface is designed to feature an intuitive layout with chat bubbles for user queries and system-generated responses. This setup allows users to seamlessly upload documents, view summaries, and interact with the system to ask legal questions.

This structured approach ensures the system effectively supports users in understanding and navigating complex legal documents, making them more accessible to non-specialists.

7 IMPLEMENTATION

The application was implemented using Python, with Streamlit as the host and the frontend for user interaction. Key libraries include PyPDF2 for text extraction, LangChain for query handling and knowledge base creation, and Hugging Face's transformers for NLP-based classification. The implementation begins by first creating a knowledgebase using LangChain and OpenAI Embeddings to incorporate Retrieval-Augmented Generation (RAG) for context-aware responses. The user is them allowed to upload a legal document PDF for queries, validated against legal-specific criteria using a pre-trained Legal BERT classification model.

The response generation system combines two components, a retriever and a generator. The retriever fetches relevant chunks of text from a locally stored, vectorized knowledge base stored in FAISS. The retrieved chunks provide context for the language model, which generates answers using OpenAI's language model, GPT 3.5 in this case. This approach ensures that responses are not only generated from the language model's pre-trained knowledge but also based on the specific legal documents and context uploaded by the user and referenced from the knowledgebase, significantly improving relevance and accuracy.

Additionally, summarization is performed using OpenAI's language model, while embeddings are generated via the OpenAI Embeddings API to build a searchable vector database. User queries are processed through a customized prompt template, ensuring responses are contextually accurate and elaborate. Moreover, the interface displays a summary of the document and maintains chat history with styled bubbles for user convenience. The system is designed to handle tokenization and classification confidence thresholds to enhance output reliability.

8 EVALUATION

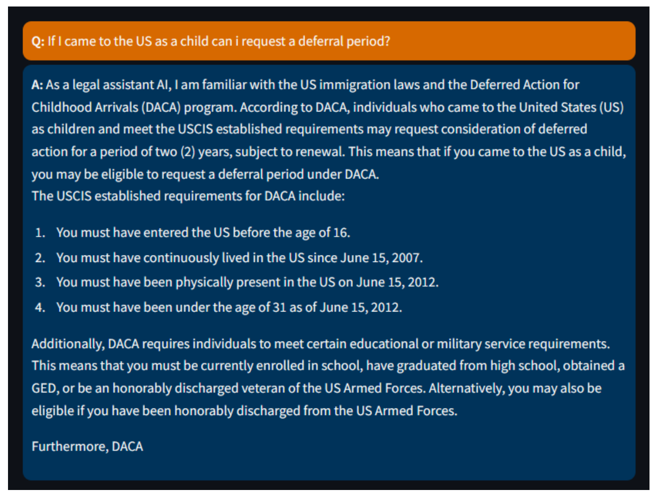

Figure 1: Finetuned Legal LLM (GPT 3.5)

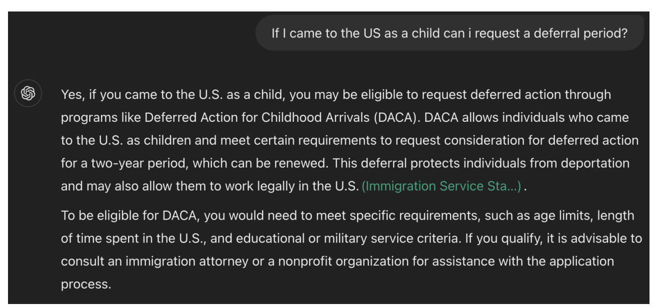

Figure 2: GPT 4.o Mini

Figures 1 and 2 present a comparison between the fine-tuned Legal LLM and ChatGPT 4.0 Mini. When queried, "If I came to the US as a child, can I request a deferral period?", the response from the Legal LLM was notably more detailed, despite operating on a smaller model (GPT-3.5), compared to the response generated by GPT-4.0 Mini. This highlights the effectiveness of the fine-tuning process in improving the ability of the legal refined model to generate detailed and contextually accurate legal responses.

In addition to offering a more detailed response, the legal refined model also includes citations for each relevant clause, enhancing the transparency and accuracy of the information provided. This feature allows for a more structured and comprehensive response, ensuring that each piece of legal information is precisely located within the document. While both GPT and the legal refined model offer similar support, the legal refined LLM system's outputs are notably more structured, with in-depth legal explanations that enhance the user experience. Furthermore, as more data is integrated and the model is refined, the clarity and precision of these responses will continue to improve, solidifying the Legal refined model's position as a superior tool for legal document analysis and query resolution.

9 EXPERIMENT/CASE STUDIES

The application was tailored within the immigration legal department to address queries pertaining to immigration law, reflecting an optimal scenario in which the model is attuned to various legal aspects. For instance, a sample legal document related to immigration from the state of California was uploaded to conduct user inquiries. The model successfully verified the document's legal validity and proceeded to generate a summary, thus facilitating subsequent user-oriented inquiries. When posed with questions specific to the document and immigration, the model demonstrated an ability to provide comprehensive responses by leveraging both its immigration-related knowledge base and its intrinsic knowledge. Furthermore, a non-legal document was introduced to ascertain the model's ability to discern between legitimate legal documents and others. In this instance, the model correctly deemed the document non-legal. Despite occasional inconsistencies, the model predominantly succeeds in distinguishing between valid and invalid legal documents. These instances illustrate the system's applicability in classifying legal documents and delivering responses that are pertinent and contextually informed.

10 DISCUSSION

The integration of NLP and vector databases demonstrates the system's capacity to efficiently handle complex legal documents. By using GPT-3.5 instead of GPT-4, we achieve a more balanced trade-off between computational efficiency and response quality while maintaining robust classification through downstream BERT training for better legal document classification. This modular design ensures scalability, with each component capable of independently enhancing overall system performance. Challenges persist in managing computational costs, particularly with large documents and extensive embeddings. Fine-tuning the model for domain-specific contexts, such as intellectual property law, alongside continued prompt engineering for question-answering, will enhance the model's precision and legal relevance. Future developments may include multi-document analysis for cross-referencing and integrating case law databases, further strengthening the system's decision-making capabilities.

11 FUTURE WORK

Future enhancements will focus on improving the system's functionality and expanding its capabilities to address more complex legal scenarios. Key directions include:

Training the Legal BERT Model for Downstream Tasks: Expanding the capabilities of the Legal BERT model through training on downstream tasks will enhance its classification accuracy, particularly in document validation and legal clause identification. This improvement will bolster the tool's ability to assess legal documents with greater precision.

Refining Responses with GPT-4: Upgrading the system to utilize GPT-4 instead of GPT-3.5 will allow for more nuanced and contextually accurate responses. By crafting refined prompts tailored for legal contexts, the system can deliver specialized and high-quality responses that better meet user needs.

Expanding the Knowledge Base: Adding more comprehensive and domain-specific data to the system's knowledge base will improve its ability to provide relevant and specialized answers. This expansion will support the analysis of a wider range of legal documents and scenarios, increasing the system's utility and reliability.

These advancements will ensure the system remains a cutting-edge solution, capable of meeting the evolving demands of legal professionals and organizations.

12 CONCLUSION

This project demonstrates the potential of leveraging advanced AI techniques to streamline legal document analysis. By integrating robust natural language processing components for text summarization and question-answering, the system addresses critical inefficiencies in legal workflows, such as document review and query resolution. Its tailored approach ensures contextual accuracy and relevance, providing a valuable tool for legal professionals and non-experts alike. The emphasis on domain-specific fine-tuning, error mitigation, and compliance ensures the system aligns with the stringent demands of the legal field. Overall, this work serves as a foundation for developing more accessible, reliable, and efficient AI-powered legal solutions for the future.

REFERENCES

[1] Inyoung Cheong, King Xia, K. J. Kevin Feng, Quan Ze Chen, and Amy X. Zhang. 2024. (A)I Am Not a Lawyer, But...: Engaging Legal Experts towards Responsible LLM Policies for Legal Advice.

[2] Jiaxi Cui, Munan Ning, Zongjian Li, Bohua Chen, Yang Yan, Hao Li, Bin Ling, Yonghong Tian, and Li Yuan. 2024. Chatlaw: A Multi-Agent Collaborative Legal Assistant with Knowledge Graph Enhanced Mixture-of-Experts Large Language Model.

[3] Xukun Liu David Demeter Mengnan Du Yongfeng Zhang Dong Shu, Haoran Zhao. 2024. LawLLM: Law Large Language Model for the US Legal System.

[4] Hang Jiang, Xiajie Zhang, Robert Mahari, Daniel Kessler, Eric Ma, Tal August, Irene Li, Alex Pentland, Yoon Kim, Deb Roy, and Jad Kabbara. 2024. Leveraging Large Language Models for Learning Complex Legal Concepts through Storytelling. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Lun-Wei Ku, Andre Martins, and Vivek Srikumar (Eds.). Association for Computational Linguistics, Bangkok, Thailand.

[5] Florian Leiser, Sven Eckhardt, Valentin Leuthe, Merlin Knaeble, Alexander Mädche, Gerhard Schwabe, and Ali Sunyaev. 2024. HILL: A Hallucination Identifier for Large Language Models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 482.

[6] Athul S, Anshul Saxena, Jayant Mahajan, Lekha Panikulangara, Shruti Kulkarni, and Dr Sanjay Bang. 2024. LegalMind System and the LLM-based Legal Judgment Query System. In *2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies*.