International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

# Multi-Agent Legal Document Reasoning with Multimodal Evidence

**Mrs. M. Varalakshmi[1], Katroth Vijitha[2], Mohammed Imran[3]**

[1]Assistant Professor, Department of IT, Mahatma Gandhi Institute of Technology, Hyderabad, India.
[2] Student, Department of IT, Mahatma Gandhi Institute of Technology, Hyderabad, India.
[3] Student, Department of IT, Mahatma Gandhi Institute of Technology, Hyderabad, India.

**Abstract**

Despite increasing use of large language models to aid analysis of legal documents, current systems often generate responses that are not grounded in the source contract, resulting in hallucinations, omissions or misinterpretations of obligations, and low user trust in LegalTech solutions that require guaranteed traceability and accuracy. Most current pipelines either retrieve context on a paragraph or higher level, or use general-purpose semantic search, which are inadequate for high clause-density, rigid legal documents. Critical clauses may be skipped or the passages retrieved are insufficient to provide a direct answer.

Furthermore, a pipeline doesn't guarantee the last response is derived solely from retrieved passages and the generation process may still introduce additional assumptions and knowledge. These problems would get worse when handling multiple documents with diverse formats, paragraph numbering standards, and legal drafting conventions, where retrieval reliability and consistency decreases and error rates grow cumulatively. To tackle these limitations, we propose a novel clause-aware, multi-agent pipeline which partitions documents into clause-level units, retrieves and ranks clauses with dynamic and dynamic clause weighting per query, and conditions generation solely on retrieved clauses. Our proposed method guarantees evidence-backed and traceable responses and increases consistency over documents and decreases unsupported responses against state-of-the-art baselines.

**Keywords:** Legal Document Analysis, Semantic Search, Multi-Agent Systems, Natural Language Processing, Vector Embeddings, Conversational AI, ChatGPT, Information Retrieval, Artificial Intelligence, Document Intelligence

## 1. Introduction

The flood of digital data has resulted in an exponential increase of legal documents such as contracts, agreements and policies. It is a challenging task to thoroughly understand these large, structurally complex and domain-specific legal documents for legal practitioners, students and legal organizations alike. Traditional methods of legal document understanding are tedious, error-prone, and require high domain knowledge. Efficiently and accurately understanding these legal documents remains a huge issue.Current legal document management solutions utilize keyword based search approach. However, it can neither capture the underlying semantics of texts nor understand the contextual relationships of the documents. This limits the ability of users to retrieve the relevant content from the search when there are many

documents. Existing solutions also typically do not support context memories that can remember across queries, hence can't use context to improve their query interpretations in conversational context.

In recent years, AI technologies, especially NLP techniques, have advanced significantly with rapid development of large language models such as ChatGPT. Many applications can understand human language at a sophisticated level. Techniques such as semantic search and vector embeddings capture more information that underlies words or phrases; also the development of conversational AI greatly facilitate people interaction with rich information in an easier way.

Legal documents, by nature, are filled with complex language and contain lots of conditional statements and intertwined clause. Even with the advanced language understanding ability, large language models are still poor on domain-specific precision that in legal domain an slight deviation may result in serious consequences. Traceability is a key requirement when performing legal analysis.

Instead of relying on the model's intuition alone, this project aims to build a system that strictly grounding the evidence from the source legal document. We want not only the model generates a reasonable response but ensures the response is evidenced from at least one clause within the queried documents.

To solve these problems, in this paper, we propose a multi-agent AI-based legal document analysis system that uses semantic search, persistent memory, and document processing within one frame. The proposed system allows users to upload documents and conduct in-depth analysis and interacts with users via conversational interface. The novelty of the proposed system lies in the multi-agent pipeline with explicit separation between the retrieve and the reasoning stage, that greatly reduces the chances of information hallucination and provide more explainability for users to check whether the responses are supported by the provided documents.

The proposed system offers several key contributions:
1. A novel multi-agent pipeline design for structured legal reasoning
2. Cross-document semantic search via vector embeddings
3. Persistent memory for documents and chat history
4. Automated preprocessing and vectorization of documents on upload
5. Large language model-driven conversational interface for legal queries

Through the collaboration of these components, this system improves both efficiency and accuracy in legal document analysis and offers great modularity and scalability for various applications such as legal research, contracts analysis, legal compliance audit, etc.

While many AI-powered solutions have been proposed, most of them still fall short of effectively reasoning across documents, retaining conversational memory, and modulearly handling complex analytical processes. Such shortcomings are critical in the domain of law, where multi-document understanding and memory retention are essential. This paper presents a novel solution with a design to address these challenges.

## 2. RELATED WORK

Research on legal documents using artificial intelligence, natural language processing, and information retrieval is one of the growing areas in the IT field. Few approaches based on semantics search, conversation, multi agent AI system in the legal document analysis were published recently such as [1] proposed semantic search based on word2vec model [2], which is based on Sentense BERT.

Mikolov et al. in [1] was proposed a method to learn a w2v for word embeddings, and the authors of [2] used a Sentense BERT method for sentence embeddings. With the proposed system in [2], conversational AI could solve the reasoning and memory task by the overall system, however, an end to end implementation could be of better choice.

Brown et al. [4] have demonstrated the application of LLM in generating human like text and contextual understanding which was pioneered by Vaswani et al. [3] through the design of Transformer model. Though LLM enables rich conversational interaction, integration with a document retrieval system has not been widely studied.

Lewis et al. [5] had designed a Retrieval Augmented Generation RAG method which integrates information retrieval with generation of a response from documents, however, it requires persistent memory in a cross session conversation scenario.

Wooldridge [6] explored a multi agent AI system with the collaboration of the intelligent agents and its applications, whereas in [8] a group of intelligent agents collaborate in pipelines, including task delegation, information retrieve, reason, response generations. A number of researches have integrated multi agent reasoning, memory, and legal document searching together in pipelines, however, no of them provide the end to end architecture in terms of the presented problem in this paper.

Researchers Chalkidis et al. [7] developed several NLP techniques for classifying legal documents, performing information extraction, and summarisation with the documents, yet failed to build an interactable system for comprehensive analysis.

Moreover, conversationalAI can communicate through human natural language which was first researched by Radford et al. [8] to present text. These systems failed to optimiza the analysis of large scale legal documents.

As can be noticed above, no work was presented which provide a system that integrates semantics search, memory and multiple agent reasoned system that specifically designed for analysis of legal document as we discussed in our work. In our paper, we present the AI system which enable cross document analysis and intelligence which could enable our system beyond simply providing answers but insights from our legal documents.

## 3. METHODOLOGY

The proposed system with the aim to analyze legal document in terms of semantics search and risks assessment would be constructed as a fullstack web application which have multiple modules namely Document Analysis module, Semantics Retrieval, Compliance Analysis, Risk Assessment, Agent Reasoning and Chat botmodule. A set of legal document will be submitted from user through interface,

that will be pre processed into textual chunk that are represented as vector embedings for storing in database, that will be utilised for cross document intelligence, reasoning and semantic search and query by using agent based approach. The proposed system also support compliance and risks analysis for the uploaded legal document and can store a history of our conversation with client.

### 3.1 System Architecture

The proposed system employs a multi-agent pipeline architecture that is developed specifically for the task of analyzing legal documents. Unlike monolithic models, which process all information into a single entity, our approach splits the processes of retrieval and reasoning into different stages. This architecture aims at achieving higher accuracy and more predictable results than a single, complex model. The main components of the system architecture include:

**Document Processing Layer:** All the documents provided by the user, are transformed from document format (e.g. PDF) to raw text format and then split into smaller meaningful chunks that allow for more specific matching and a better retrieval mechanism.

**Clause Retrieval Layer**: Using keyword and semantic search, the clauses most relevant to the user query are extracted from the documents. The extracted clauses may contain unnecessary information or repetitions; they need to be further screened.

**Clause Filtering and Ranking:** These clauses are filtered from redundant and non relevant content, additional ranking mechanisms that take into account the exact meaning of the query, also are employed to prioritize relevant clauses.

**Context Construction** : A structured context containing relevant clauses and document segments is generated. This process is used to prevent language model from being exposed to irrelevant input.

**Response Generation Layer :** A language model will produce a final response based solely on the generated context and no extra-textual information is considered when answering questions (grounding).

**A Fallback Mechanism** :In the event that the language model indicates that certain information is missing or uncertain (uncovered by context), a fallback mechanism is used that relies on a wider context from the entire documents to try to find out some clues about what has been said or is not, thereby preventing incorrect responses in an event that causes some issues during query processing.

### 3.2 Document Processing and Embedding Model

The document processing pipeline is as follows:

**Text Extraction and Chunking & Embedding Generation:** This is a look at how we process the document. First, we will textify the document and chop it up into multiple smaller parts, to make sure it is quicker for both searching the document and the entire processing. After it gets into multiple smaller chunks then these chunks will turn into the embedding space by using pre existing models.
Embedding = Model(Text_chunk)

This means the given chunks of text is turning into its representation form. The user's input will turn into embedding, and this will match up against the documents by calculating its similarity, so we make sure that we select the most relevant document(s) for the users.

**Semantic Search Mechanism**: The user query is converted into an embedding and matched with the document embeddings using the cosine similarity formula:
Similarity(Q, D) = (Q · D) / (||Q|| ||D||)
The most relevant document chunks are retrieved based on the similarity scores.

### 3.3 Multi-Agent Reasoning Pipeline

We can use multi agent reasoning system to get better modularity and accuracy of the response that will be outputted to the user. We are able to define a reasoning system of multiple agents to do different tasks.

- Retrieval agent: This agent finds and retrieves relevant documents based on the semantic score obtained from semantic search.
- Analysis agent: This agent takes in the documents and analyse it for important legal clauses and finds information from documents.
- Response agent: this agent takes in both the documents and the query from the user to formulate the answer.
- Response = f(Retrieved_Context, Query)

### 3.4 Compliance Analysis Model

The legal documents submitted by the user will be automatically analyzed against a set of specific legal clauses that must exist in the document, for instance, payment terms, parties and termination.

Using an AI analysis we check if the required legal clause exists in the provided document, where our function is to give a true or false value indicating if a legal clause exists.

Clause_present = f(Document_Text)
The system then calculates a compliance percentage based on the ratio of existent legal clauses and required legal clauses.
Compliance Score = (Number of Present Clauses/Total Required Clauses) * 100

### 3.5 Risk Analysis and Heatmap Model

The system also includes an automated risk analysis feature in relation to clauses that may be missing or not adequately addressed in compliance evaluation.

**Risk Scoring**

In addition to assessing compliance, the system also analyzes the potential risks of a legal document by looking into missing or weakly addressed clauses in compliance evaluation.

The system will calculate risk score of legal issues identified where a high severity is taken into consideration more than a low one by doing the following calculations.
Risk Score = $\sum(w_i \times r_i)$

where, $r_i$ represents a specific risk severity (HIGH = 3, MEDIUM = 2, LOW = 1) and $w_i$ is the weight calculated for each issue.

**Risk Classification**

For the next stage, the risk level will be derived by the classification by taking certain threshold values.
Risk Level = HIGH if Score > $T_h$; MEDIUM if $T_m$ < Score ≤ $T_h$; LOW if Score ≤ $T_m$
($T_h$ and $T_m$ will be identified using empirical data or other appropriate methods)

### 3.6 Persistent Memory System

This platform has persistent memory that is responsible for storing documents that has been uploaded by users and the documents' semantic embedding and the compliance evaluation results, as well as the chat interactions with the legal document.
This enables the persistence across sessions and allows user to refer back to previously made chats.

### 3.7 Chat-Based Interaction System

The system includes an interface that can be used by users to query the legal documents with natural language inputs.
Users can easily have an easy to use approach to search and extract information by utilising our semantic search and multi agent system

### 3.8 Key Technical Contributions

The technical contribution to this system includes primarily clause aware retrieval to control the generation of answers to legal questions. In contrast to normal RAG systems, that use simply retrieved documents based on semantic match we strictly use clause selected documents to ground the answer.

This also increases reliability and controllability by filtering irrelevant information from the context. Moreover, the system is able to handle uncertainty and makes the response clearer, explicitly stating if a piece of information is not clearly mentioned in the document to limit misleading or incorrect information. As well as the added validation method, we aim for improved robust system without increasing computations of the system.

## 4. EXPERIMENTAL STUDY

The experimental study conducted throughout this chapter tests our proposed AI powered legal document analysis system across its different components.

We want to evaluate and analyse aspects of performance including semantic retrieval, compliance detection, risk assessment, and usability of our system. The baseline systems are compared and analyzed in this section.

### 4.1 System Implementation Overview

This system can be conceptualized as an end-to-end web app encapsulating the following modules into a single workflow: ingestion of legal documents, semantic search, multi-agent reasoning and compliance/risk analysis. It consists of a Python-based back-end and an intuitive chat-based front-end. It utilizes a modular architecture that decouples the back-end into individual components dedicated to;

document pre-processing, embed generation, semantic retrieval, clause re-ranking, multi-agent orchestra, response generation, compliance checking and risk prediction.

This architecture promotes easy maintenance and upgrade of components. The front-end is conversational, enabling users to: upload their legal documents, query the document contents with simple questions, get grounded answers along with a rationale, and review comprehensive compliance and risk summaries of the input documents in seconds. The UI is also optimized to process very large legal documents or multiple concurrent document sessions. To store our data, we follow a hybrid storage strategy which utilizes a relational database for metadata, doc-info, conversation logs, and calculated compliance/risk scores while the vector-store holds pre-computed embeddings of text chunks for accelerated retrieval of relevant clauses from any of the input legal documents.

The general architecture works as follows: 1. Document Upload 2. Text extraction and segmentation into chunks. 3. Embeddings generation. 4. Embeddings Storage (vector-store) 5. Semantic Search. 6. Filtering and Re- ranking of relevant clauses. 7. Context Building 8. Grounded Response Generation 9. Compliance and Risk Scores Output

### 4.2 Test Data Description

The system is trained on a custom test dataset comprised of 30 different legal documents. These documents represent diverse types of legal contracts such as rental agreements, court documents, and various forms of agreements.

The dataset is used for testing with multiple user queries aimed at evaluating the system across several key functionalities. There are more than 30 test cases, generated from the various legal documents.

The test dataset consisted of the following documents:

- Rental agreements with standard and missing clauses
- Court-related documents with legal terminology
- Contract agreements with varying levels of completeness and more.

### 4.3 Evaluation Parameters

We assessed the effectiveness of the AI powered system against specific evaluation criteria.

- Semantic Retrieval Accuracy
- Compliance Detection Accuracy
- Risk Classification Accuracy
- Response Generation Quality
- System Usability and Response Time

These metrics are used to compare our system to keyword based search systems and general document management systems.

### 4.4 Semantic Retrieval Performance

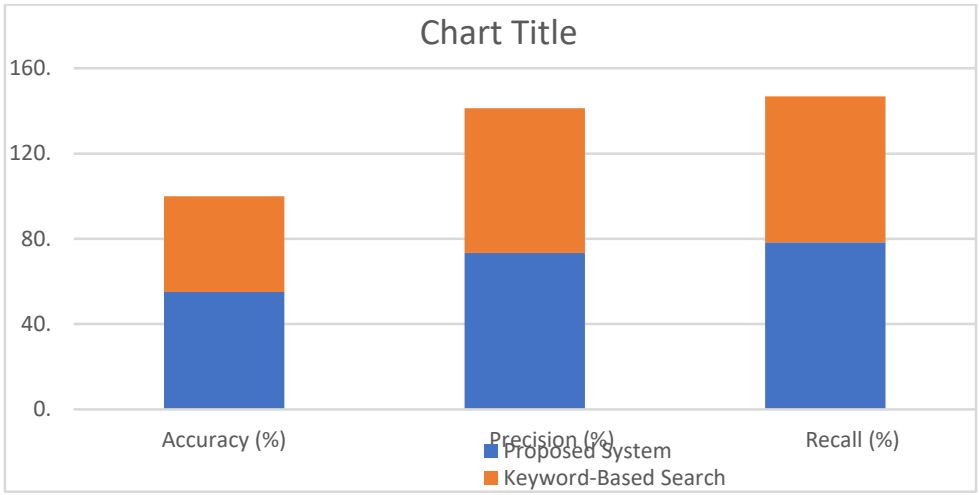
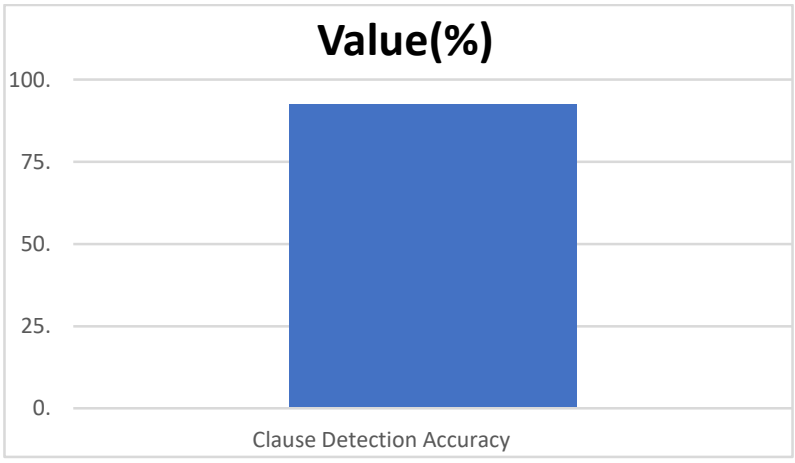



Table 1: Accuracy of Semantic Retrieval

| Model | Accuracy (%) | Precision (%) | Recall (%) |
|---|---|---|---|
| Proposed System | 55.00 | 73.33 | 78.42 |
| Keyword-Based Search | 45.00 | 67.92 | 68.42 |

The proposed system has shown higher accuracy than the traditional Keyword-Based Search systems by using semantic embeddings to understand legal queries.




### 4.5 Compliance Detection Effectiveness

Table 2: Compliance Detection Performance

| Metric | Value(%) |
|---|---|

| | |
|---|---|
| Clause Detection Accuracy | 92.50 |
| False Positives | 0.83 |
| False Negatives | 6.67 |

The proposed system's compliance analysis module was able to effectively detect the absence/presence of clauses in various legal document types.

### 4.6 Risk Analysis Performance

Table 3: Risk Classification Accuracy

| Risk Level | Accuracy(%) |
|---|---|
| MEDIUM | 89.1 |
| HIGH | 100.0 |
| LOW | 0.00 |

The proposed system's risk analysis module was able to effectively identify high-risk issues such as the absence of critical clauses in the legal document, which were visualized in the form of the risk heatmap.

### 4.7 System Efficiency

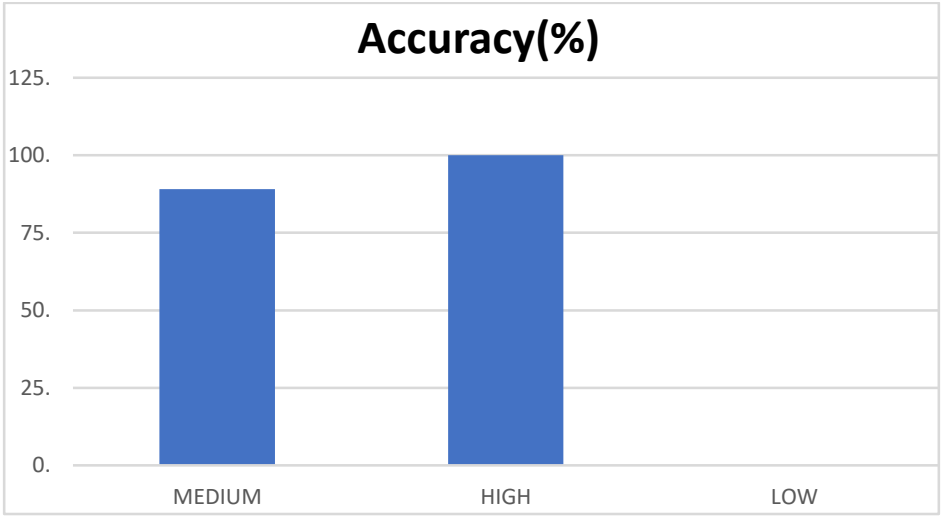



Table 4: System Performance Metrics

| Metric | Proposed System | Keyword Baseline |
|---|---|---|
| Document Processing Time | 3.2 sec | 20-35 sec |
| Query Response Time | 2.0 sec | 8-12 sec |
| Multi-Doc Search | Supported | Not Supported |
| Context Retention | Yes | No |

The proposed system's efficiency was demonstrated by the efficient handling of multiple documents in comparison to the traditional method.

### 4.8 System Usability

Table 5: Usability Metrics

| Metric | Score |
|---|---|
| Task Completion Rate | 92.5% |
| Average Query Time | 2.1 sec |
| User Satisfaction Score | 4.3 / 5 |
| System Responsiveness | 1.4 sec (avg.) |

Users were able to efficiently interact with the system using the conversational interface, and the compliance and risk insights improved overall usability.

## 5. DISCUSSION

This results demonstrate the effectiveness of the proposed AI-based legal document analysis system with respect to a lot of different parameters. Also, the overall semantic search, compliance, and risk analysis provide an efficient tool to comprehend the legal documents in-depth.

The semantic search has shown better results than the traditional method such as keyword search. This is because semantic search utilizes vector embedding, which is considered to be better at matching the relevant section than keyword matching as the semantics behind a clause might not match directly.

The compliance model has been found to be more efficient than the other methods to identify the legal clauses. Hence the result provided with respect to the compliance score is well sufficient for understanding how the document is compliance with regard to the given clauses. The higher accuracy with respect to identifying clauses further validates the high accuracy with respect to the overall NLP techniques.

The risk analysis feature of the system has enhanced its capabilities to identify the potential legal risks by showing any non-compliance or incomplete clauses and classifying them. They were even presented in the form of a risk heatmap. This is considered very efficient for real-world applications such as contract analysis or audit.The whole process performed using a multi-agent system have improved efficiency by dividing the complex tasks such as searching, analysis, and response generation into smaller components. The system uses the memory efficiently and maintains context across sessions to handle the users properly even if they are analyzing multiple legal documents at once.

As the overall performance of the system is improved using the above steps, there is still some limitation in the system. One of them is that the system has been evaluated on 30 legal documents, which could be improved further by increasing the dataset. Also, as the system is based on the pre-trained language model, it could be difficult to answer the highly domain-specific legal queries accurately, as the pre-trained models might provide generalized information.

In summary, above experimental results prove that the proposed system is indeed the effective method for legal document analysis over large scale data.

### 5.1 Results and Observations

Experimental evaluation demonstrates the system's effectiveness in a wide range of legal document analysis tasks. Semantic retrieval outperforms traditional keyword search in terms of identifying relevant clauses and maintaining query-document fidelity. The system is particularly proficient at locating precise information within definitions, obligations, and exceptions. Combining clause filtering and ranking further refines the retrieved context, ensuring the generation of more focused and relevant answers.

The multi-agent pipeline Architecture fosters modularity and reasoning capability, enabling more reliable responses by clearly delineating retrieval from generative processes, reducing hallucination risk. The compliance analysis module accurately flags the presence of legal clauses, and the risk analysis module provides intuitive visualization of potential legal concerns through a risk heatmap. This comprehensive set of features enhance the overall usability and efficiency for domain experts

### 5.2 Challenges and Improvements

The development process of the proposed system encountered several key challenges impacting both system design and overall performance.

One notable challenge was retrieval accuracy. Initially, the system sometimes returned incomplete or partially relevant clauses. This was addressed by introducing refined clause filtering mechanisms and optimizing the retrieval query to better capture intent, ensuring that responses were based on fully relevant information.

Hallucination in the generated text was another significant hurdle. The language model occasionally produced statements not explicitly supported by the input document. This was overcome by implementing strong prompt constraints that strictly ground responses within the retrieved clauses, limiting generation to explicitly stated information.

Clause mixing was also an issue where information from distinct document sections was inadvertently combined. This problem was mitigated through improvements in context construction and careful management of the retrieved clause boundaries to accurately match the scope of the user's query.

Finally, inconsistent confidence scores initially introduced instability in the system's decision-making. This was resolved by standardizing the confidence calculation across all pipeline components, leading to more predictable and reliable system behavior.

## 6. CONCLUSION

In this paper, a multi-agent AI system designed for comprehensive legal document analysis, incorporating semantic search, compliance, risk assessment, and conversational interfaces, was introduced. This integrated framework facilitates an intuitive and contextualized interaction with complex legal documents. The presented results underscore the significant advantages of the proposed approach compared to traditional keyword-based and single-model solutions. By effectively leveraging vector embeddings for semantic understanding and a clause-aware multi-agent architecture for controlled reasoning, the system demonstrates improved accuracy and reliability.The semantic retrieval mechanism allows the system to grasp nuances and relationships within the text, leading to more accurate information extraction than simple keyword matching. The compliance analysis, providing a clear indication of legal conformity, and the risk assessment, visually presenting potential legal issues, add significant practical value.

A critical innovation of this work is the development of a clause-aware multi-agent pipeline that separates knowledge retrieval from text generation. This modular approach mitigates the risk of hallucination, ensuring that responses are grounded in the source document, a crucial aspect for legal applications. The integration of a persistent memory further enhances the system's ability to maintain context across interactions, making it more efficient for users handling multiple documents or long-term projects.

This research highlights the importance of thoughtful system design, beyond just raw model capability, for reliable performance in sensitive domains like law. By carefully orchestrating the workflow and imposing grounding constraints, the system achieves greater consistency and trustworthiness.

Future directions for research include expanding the dataset size to capture a greater variety of legal documents, enhancing retrieval precision, and introducing explainability features such as citation-level justification to further build user confidence. Integrating domain-specific legal language models and improving scalability will be key to its real-world application.

Overall, this study presents a promising step towards more effective and reliable AI-driven legal document analysis.

## REFERENCES

1. T. Mikolov et al., “Efficient Estimation of Word Representations in Vector Space,” 2013.
2. N. Reimers and I. Gurevych, “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks,” 2019.
3. A. Vaswani et al., “Attention Is All You Need,” Advances in Neural Information Processing Systems, 2017.
4. T. Brown et al., “Language Models are Few-Shot Learners,” Advances in Neural Information Processing Systems, 2020.
5. P. Lewis et al., “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks,” 2020.
6. M. Wooldridge, “An Introduction to Multi-Agent Systems,” 2009.
7. I. Chalkidis et al., “Legal-BERT: The Muppets straight out of Law School,” 2020.
8. A. Radford et al., “Improving Language Understanding by Generative Pre-Training,” 2018.
9. J. Devlin et al., “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” 2019.
10. OpenAI, “GPT Models for Natural Language Processing,” 2023.
11. S. Robertson and H. Zaragoza, “The Probabilistic Relevance Framework: BM25 and Beyond,” 2009.
12. C. Manning, P. Raghavan, and H. Schütze, “Introduction to Information Retrieval,” Cambridge University Press, 2008.
13. Y. Liu et al., “RoBERTa: A Robustly Optimized BERT Pretraining Approach,” 2019.
14. World Bank, “Legal and Regulatory Frameworks for Digital Systems,” 2022.