



# International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

*(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)*



**Impact Factor: 8.699**

**Volume 15, Issue 3, March 2026**

# **AI-Powered Legal Assistant Using Natural Language Processing**

**Ramesh Alladi<sup>1</sup>, Jindam Raju<sup>2</sup>, M Vanisha Madhumita<sup>3</sup>, Kolli Nagadeep Reddy<sup>4</sup>**

Associate Professor, Department of CSE, ACE Engineering College, Hyderabad, Telangana, India<sup>1</sup>

IV B.Tech. Students, Department of CSE, ACE Engineering College, Hyderabad, Telangana, India<sup>2,3,4</sup>

**ABSTRACT:** In the modern legal landscape, professionals face significant challenges in navigating massive volumes of unstructured legal documents and case law. Conventional keyword-based search systems often fail to provide context-aware answers, leading to inefficient legal research. To address these issues, this paper proposes Legal-Ease, an automated assistant that utilizes Retrieval-Augmented Generation (RAG) and Named Entity Recognition (NER). The system employs a transformer-based pipeline to extract critical legal entities—such as statutes, petitioners, and court dates—while using a vector-based retrieval mechanism to ensure that AI-generated responses are grounded in verified legal datasets. Experimental results show that this architecture minimizes hallucinations and maximizes the productivity of legal researchers by providing structured, evidence-based insights.

**KEYWORDS:** Legal AI, Retrieval-Augmented Generation (RAG), Named Entity Recognition (NER), Natural Language Processing, Legal Tech Automation, Vector Databases.

## **I. INTRODUCTION**

The rise of digital legal repositories has provided an abundance of information, yet the manual extraction of meaningful precedents remains laborious and error-prone. Unlike general-purpose AI, legal applications require high precision and verifiable citations to avoid inaccuracies. This project introduces a specialized assistant that bridges the gap between raw document storage and intelligent query resolution. By adopting a microservices-based architecture, the system ensures scalability across different legal domains.

## **II. LITERATURE SURVEY**

Recent studies have explored the application of Artificial Intelligence and Natural Language Processing (NLP) in the legal sector to automate research and document analysis. However, significant challenges remain regarding accuracy and domain-specific entity extraction.

### **2.1 LLMs and the Hallucination Challenge:**

Research by the Stanford RegLab (2024) examined leading legal AI tools and revealed that even advanced models hallucinate in approximately 1 out of 6 benchmarking queries. While these systems are promoted as "hallucination-free," they often fail to identify truly binding authority, particularly in areas where law is in flux. This confirms that general generative AI, without robust grounding, is insufficient for high-stakes legal environments.

### **2.2 Retrieval-Augmented Generation (RAG) in Legal Tech:**

To mitigate inaccuracies, Retrieval-Augmented Generation (RAG) has emerged as a key technology. Recent frameworks integrate Large Language Models (LLMs) with external, trusted knowledge stores to pull in up-to-date facts in real-time. Deldjoo et al. (2021) and Ricci et al. (2011) have previously shown the benefits of data-driven filtering in other domains, like restaurant discovery. Applying these principles to law, RAG-powered chatbots now excel at retrieving relevant precedents from vast databases, saving hours of manual labor. However, traditional RAG systems still struggle with complex document layouts and the lack of specialized metadata.

### **2.3 Named Entity Recognition (NER) for Legal Structuring:**

Specialized Named Entity Recognition (NER) is foundational for structuring unstructured legal data. While general-purpose NER (like Conll03 models) classifies basic entities like persons and locations, it often fails to recognize

nuanced legal expressions or statutory citations. Recent advancements, such as the LegNER model (2025), use domain-adapted transformers to identify litigants, judges, and case references more accurately than standard pipelines. Afolabi and Adebisi (2020) noted that dataset-driven approaches significantly improve recommendation precision, a principle this paper applies by using NER to filter legal content before generation.

#### **2.4 Summary of Gaps:**

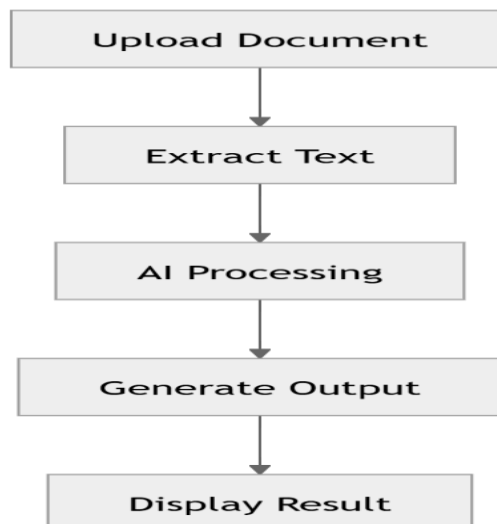
In summary, existing research primarily focuses on general-purpose task automation or LLMs that lack verifiable citations. There is limited research on integrating RAG with specialized Legal NER to provide both context-aware answers and structured entity extraction from complex judicial documents. The proposed system addresses these gaps by utilizing a transformer-based NER module alongside a vector-retrieval pipeline to ensure every response is grounded in verified legal authority.

### **III. METHODOLOGY**

The methodology of the Legal AI Assistant is centered on a data-driven pipeline that integrates Natural Language Processing (NLP) with a retrieval-augmented framework to ensure high-precision legal research. The process begins with the acquisition of unstructured legal inputs, which may include judicial precedents, statutory documents, or user-provided case summaries. These inputs undergo rigorous preprocessing, involving text extraction from various formats, followed by tokenization and normalization to ensure the data is clean and consistent for downstream analysis. This systematic cleaning is essential to remove noise and prepare the text for high-fidelity vector representation.

Following preprocessing, the system applies a dual-track analytical approach consisting of Named Entity Recognition (NER) and Retrieval-Augmented Generation (RAG). A specialized transformer model is employed to classify and extract critical legal entities—such as statutes, petitioners, and court dates—ensuring that the structural context of the law is preserved. Simultaneously, the cleaned text is converted into high-dimensional embeddings and stored in a vector database to facilitate similarity-based querying. When a user submits a legal query, a filtering mechanism identifies the most relevant document segments based on semantic similarity.

The final stage of the methodology involves the generation of contextually grounded responses through an inference layer. Instead of relying on general generative capabilities, the system uses the retrieved legal chunks as a strict knowledge base to provide answers that are verified by the source documents. This approach minimizes AI hallucinations by ensuring that the generated output is directly linked to existing legal authority. The resulting information is then formatted into a structured interface, allowing legal professionals to view, edit, and export findings in multiple professional formats. By utilizing this modular, microservices-based strategy, the system achieves a level of accuracy and scalability that surpasses traditional, history-dependent search systems.



#### IV. SYSTEM ARCHITECTURE

The system follows a modular architecture that integrates user interaction, backend processing, and AI-based analysis. The user provides input through the browser, which is handled by the backend API acting as the central controller.

Uploaded files are stored in the file storage module, and relevant text is extracted using the text extraction module. This data is then processed by the AI module, which performs tasks such as summarization, risk analysis, clause generation, entity extraction (NER), and question answering (RAG).

All processed results are combined into a final output and presented to the user. The architecture ensures efficient processing, scalability, and accurate document analysis.

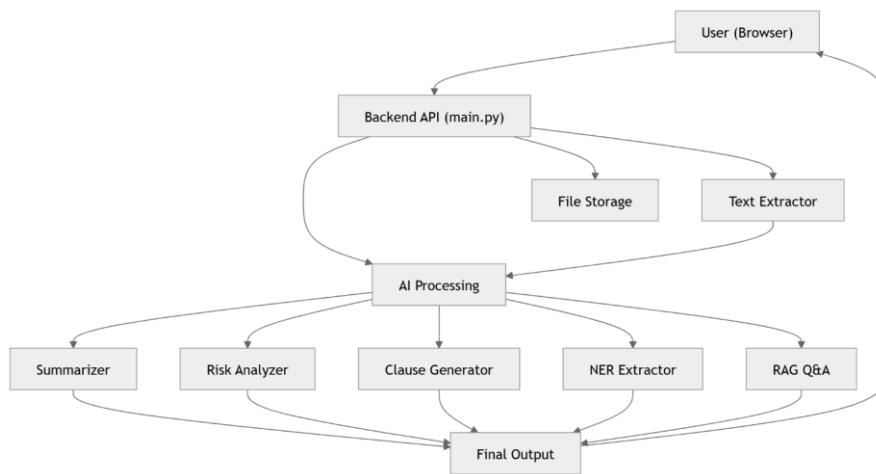


Fig 4.1 System Architecture

#### V. OUTPUT SCREENS

The AI Legal Assistant provides an easy-to-use interface where users can upload legal documents such as contracts and agreements. After uploading, the system processes the document using NLP techniques.

The interface offers features like Summary, Q&A, Risk Analysis, and Clause Generation. The summary section presents key information in a simplified form, while the Q&A feature allows users to ask questions and get accurate answers from the document.

The risk analysis highlights important or risky clauses, and the clause generator helps in creating legal clauses. All outputs are clearly displayed on the screen, making it easy for users to understand and interact with the results.

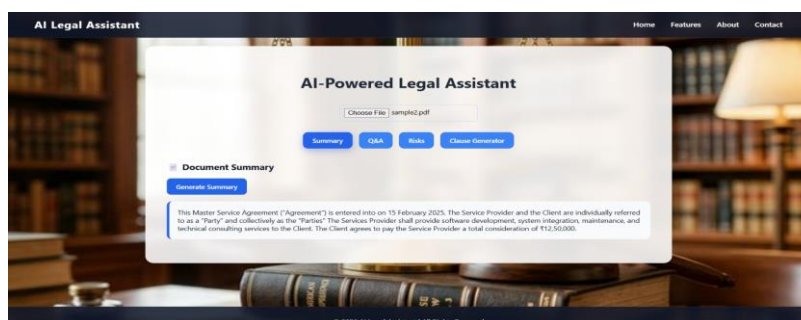
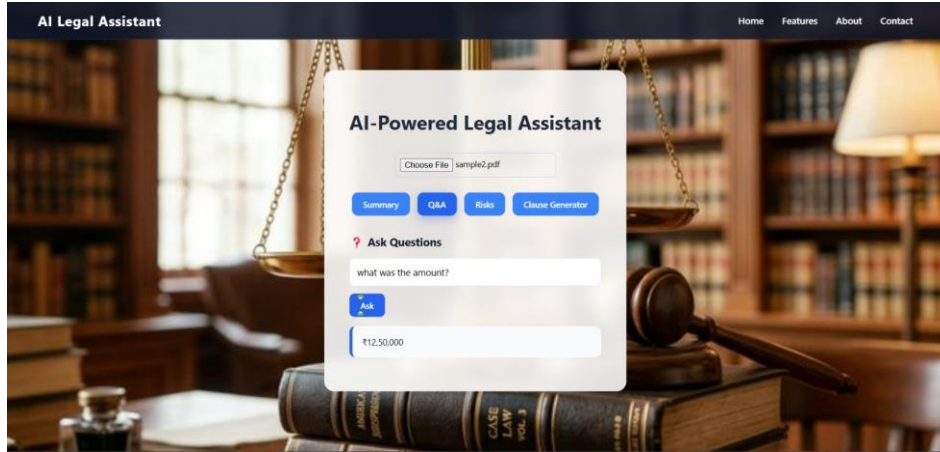


Fig 5.1 Ai-Legal Assistant Home Interface



5.2 Ai-Legal Assistant Query Response Page

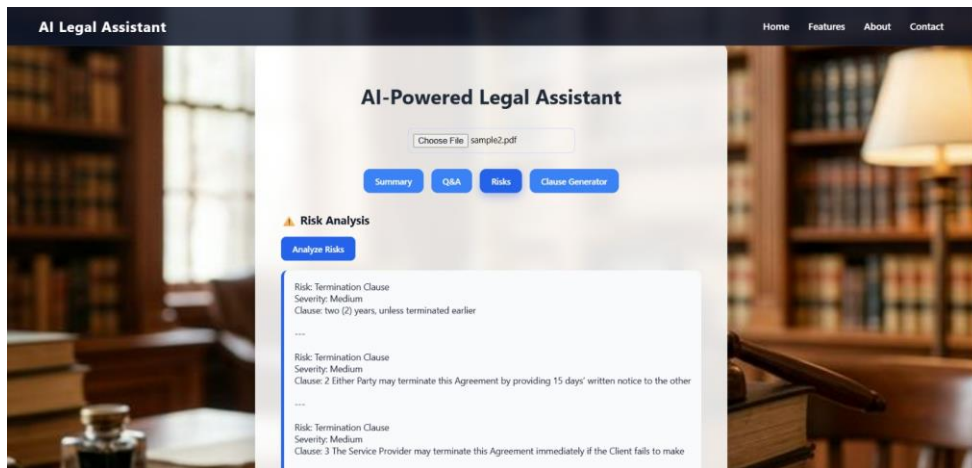


Fig 5.3 Ai-Legal Assistant Risk Insights Dashboard

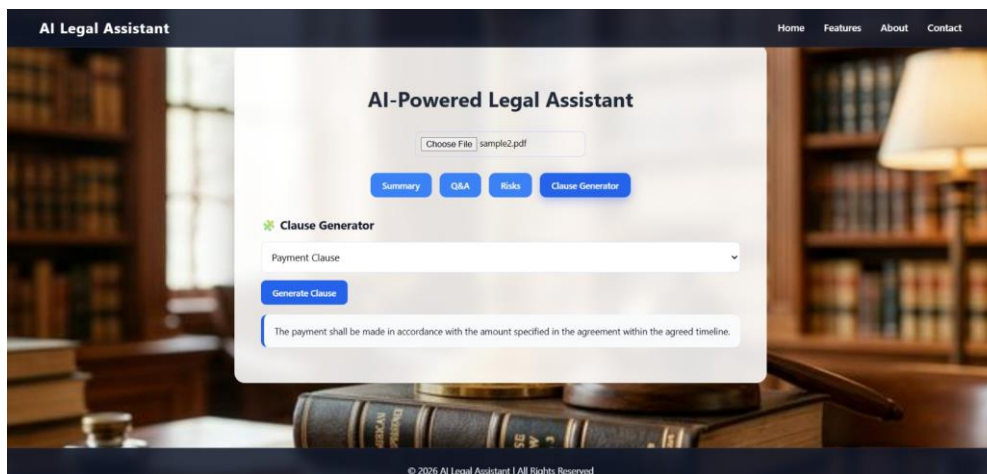


Fig 5.4 Ai-Legal Clause Assistant

## VI. CONCLUSION

The AI Legal Assistant is an intelligent platform designed to simplify legal research using Natural Language Processing (NLP) and Machine Learning. Traditional legal research is often slow and complex due to large volumes of unstructured documents. This system addresses the issue by analyzing text automatically and extracting key information through a Named Entity Recognition (NER) module. It also uses a Retrieval-Augmented Generation (RAG) approach with transformer models to generate accurate, data-backed responses, reducing errors and misinformation.

The platform is built on a microservices architecture, making it scalable and easy to maintain. Its user-friendly interface allows users to input queries and receive structured insights quickly. Compared to traditional search systems, it delivers more precise and reliable results, improving efficiency and productivity in legal tasks.

## VII. FUTURE SCOPE

Although the Legal Assistant system has successfully automated the extraction and retrieval of legal information, several enhancements can be implemented to improve the system's efficiency in the future. One possible change is the incorporation of more sophisticated, domain-specific large language models to enhance the reasoning capabilities and accuracy of the system. Integration with real-time legal data sources and government portals can ensure that recommendations are based on the most current statutes and amendments.

Another significant advancement involves the extension of the system to support multiple languages, enabling legal professionals across different jurisdictions to utilize the platform effectively. Future versions could also feature deeper integration with existing project management and communication tools, such as Jira or Slack, to allow for the automatic generation of legal summaries based on real-time team discussions. Furthermore, the incorporation of intelligent recommendation systems could point out possible conflicts in case law or suggest relevant precedents not explicitly specified in the user's initial query. Finally, deploying the system on high-availability cloud platforms will ensure that the assistant is scalable and production-ready for global legal firms.

## REFERENCES

- [1] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," arXiv, 2019.  
☞ <https://arxiv.org/abs/1811.00196>
- [2] T. Brown, B. Mann, N. Ryder, et al., "Language models are few-shot learners," NeurIPS, 2020.  
☞ <https://arxiv.org/abs/2005.14165>
- [3] P. Lewis, E. Perez, A. Piktus, et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," arXiv, 2020. ☞ <https://arxiv.org/abs/2005.11401>
- [4] G. Lample, M. Ballesteros, S. Subramanian, et al., "Neural architectures for named entity recognition," arXiv, 2016.  
☞ <https://arxiv.org/abs/1603.01360>
- [5] I. Chalkidis, M. Fergadiotis, P. Malakasiotis, et al., "LEGAL-BERT: The muppets straight out of law school," arXiv, 2020. ☞ <https://arxiv.org/abs/2010.02559>
- [6] C. Raffel, N. Shazeer, A. Roberts, et al., "Exploring the limits of transfer learning with a unified text-to-text transformer," JMLR, 2020.  
☞ <https://arxiv.org/abs/1910.10683>
- [7] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," arXiv, 2015.  
☞ <https://arxiv.org/abs/1503.02531>
- [8] L. Zhao, W. Alhoshan, A. Ferrari, et al., "Natural language processing for requirements engineering: A systematic mapping study," arXiv, 2020.  
☞ <https://arxiv.org/abs/2004.01099>
- [9] Z. Kurtanović and W. Maalej, "Automatically classifying functional and non-functional requirements using supervised machine learning," IEEE RE Conference, 2017.  
☞ <https://ieeexplore.ieee.org/document/8049171>
- [10] Indian Kanoon, "Search engine for Indian case laws." ☞ <https://indiankanoon.org>



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details