International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3639-3643
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0475

AI-Driven Legal Research Engine for Commercial Courts

Dr. Renuka Kajale [1], Prof. Neha Bhagawat [2], Sushant Kadam[3], Tanishka Kadam[4], Preeti Pingale[5]
[1,2]Assistant Professor, Computer Engineering, NMIET, Talegaon Dabhade, Maharashtra, India
[3,4,5] UG - Computer Engineering, NMIET, Talegaon Dabhade, Maharashtra, India
Emails: *renukajale@gmail.com[1], nehasbhagwat@gmail.com[2], sushantkadam530@gmail.com[3], tanishkadeepakkadam@gmail.com[4], preetipingale290@gmail.com[5]*

Abstract
The AI-Driven Legal Research Engine is designed to improve the efficiency and accuracy of legal research in commercial courts by handling large volumes of legal data such as judgments and statutes. It enables deeper understanding of legal texts beyond simple keyword searches, providing more relevant and context-aware results. The system offers concise summaries of lengthy legal documents and allows users to ask complex legal questions through an interactive interface. By automating analysis and information retrieval, it reduces manual effort and saves time. It also improves consistency and accessibility of legal information. Overall, it provides a scalable and intelligent solution for modern legal research.
Keywords: *Retrieval-Augmented Generation (RAG), Legal AI, Indian Commercial Law, Semantic Search, FAISS, DeepSeek, Groq, Failover Architecture, Explainable AI, Failover Architecture, Legal Research Automation, Commercial courts*

1. Introduction

Legal research plays a crucial role in the functioning of commercial courts, where accurate analysis of statutes and judicial precedents is essential for resolving complex disputes. However, the increasing volume of digital legal data has made traditional research methods slow and inefficient. Most existing platforms rely on keyword-based search techniques, which often fail to capture the true context and intent of legal queries. This leads to irrelevant results and increased effort for legal professionals in identifying useful information[11-13]. To overcome these limitations, the proposed system adopts an AI-driven approach that enables semantic understanding and context-aware retrieval of legal documents. It utilizes a modern full-stack architecture combined with advanced models to support intelligent search, summarization, and question answering. By incorporating a Retrieval-Augmented Generation framework along with vector-based search, the system delivers accurate and meaningful insights.

2. Mathematical Model

2.1. Semantic Search Module

2.1.1. Introduction

The Semantic Search module uses a vector space model (VSM) to represent documents as high-dimensional vectors. It employs TF-IDF (Term Frequency–Inverse Document Frequency) for feature weighting, where each legal text is transformed into a vector:

$$V_d = [W_{1,d}, W_{2,d} ,, W_{n,d}]$$

The term frequency represents local importance:

$$tf(t,d)$$

The inverse document frequency captures global specificity:

$$idf(t,D) = \log\left(\frac{N+1}{df(t)+1}\right)+1$$

2.1.2. Outcome

Relevance is determined using Cosine Similarity between the query vector V_q
and document vector V_d

$$\text{sim}\,(V_q,V_d) = \frac{V_q.V_d}{||V_q||\,||V_d||}$$

This produces a similarity score in the range:
[0,1]

2.2.Document Summarization Module

2.2.1. Introduction

The Summarization module utilizes transformer-based models (DeepSeek/Groq) to perform

Sequence-to-Sequence mapping. A legal document is treated as:

$$S = (x_1, x_2, \ldots, x_n)$$

and mapped to a summary:

$$T = (y_1, y_2, \ldots, y_m),\ m \ll n$$

2.2.2. Outcome

The objective is to maximize the conditional probability:

$$P(T \mid S) = \prod_{i=1}^{m} p(y_i \mid y_{<i}, S, \theta)$$

2.3. Legal Q&A

2.3.1. Introduction

The Legal Q&A module implements Retrieval-Augmented Generation (RAG). It uses FAISS for efficient K-Nearest Neighbor (KNN) search in vector space:

$$R^d$$

2.3.2. Outcome

Given a query q, relevant contexts are retrieved by minimizing Euclidean distance:

$$L_2 = \| V_q - V_{ci} \|$$

The final answer is generated by maximizing:

$$P(a \mid q, c, \theta)$$

where:

$$c \in \{ c_1, c_2, \ldots, c_k \}$$

2.4. Explainable AI

2.4.1. Introduction:

Explainable AI focuses on simplifying complex legal language into understandable text through semantic transformation.

2.4.2. Outcome:

Technical terms are identified by comparing distributions using Kullback-Leibler divergence:

$$D_{KL}(P \| Q)$$

where P represents the legal corpus distribution and Q represents general language distribution.

3. System Importance

The AI-Driven Legal Research Engine plays a crucial role in addressing the inefficiencies of existing legal research tools used in Indian commercial courts. Unlike traditional platforms that rely on keyword matching, this system introduces a hybrid semantic search approach using TF-IDF, cosine similarity, and FAISS-based vector retrieval, enabling more accurate identification of relevant case laws even when queries are phrased differently[1-5]. The integration of DeepSeek and Groq models allows the system to generate high-quality summaries and precise answers with significantly lower latency, making it suitable for real-time legal research scenarios. This is particularly important for legal professionals who need quick access to relevant judgments without manually reviewing lengthy documents. The system also supports document upload and automated text extraction, allowing users to analyze their own case files and obtain structured insights. This feature enhances practical usability by bridging the gap between static legal databases and dynamic case analysis. Security mechanisms such as JWT authentication, encrypted password storage using bcrypt, and Google OAuth ensure safe access and protect sensitive legal information. Additionally, the use of a scalable full-stack architecture (React, Node.js) ensures smooth user interaction and efficient backend processing shown in Figure 1.

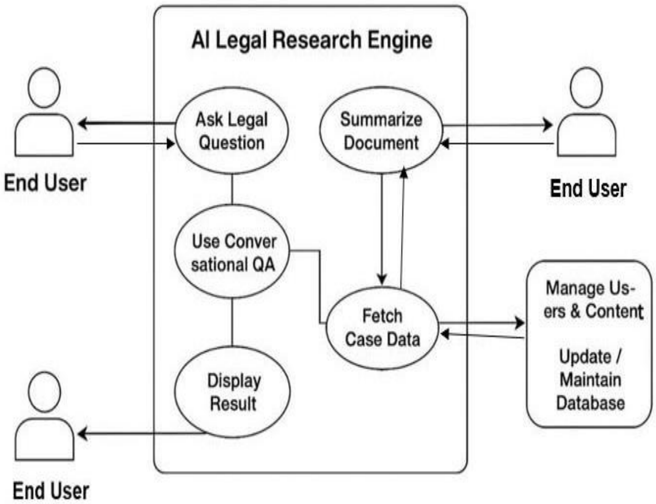

Figure 1 System workflow of the AI-Driven Legal Research Engine

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3639-3643
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0475

Table 1 Literature View

Sr No.	Year	Title	Technique	Methodology	Limitation	Paper Link
1	2024	Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools	Evaluation of LLM-based legal research tools (GPT-4, Lexis+, Westlaw)	Benchmarked various AI tools for legal reasoning, factual accuracy, and hallucination rates in real legal queries.	Existing systems often generate hallucinated citations and lack transparency and retrieval grounding	https://arxiv.org/abs/2405.20362
2	2023	SemEval 2023 Task 6: LegalEval – Understanding Legal Texts	Transformer-based models (LegalBERT, RoBERTa))	Organized a shared NLP task for legal text classification and entailment using transformer architectures.	Focused only on classification, no summarization or QA integration.	https://arxiv.org/abs/2304.09548
3	2022	Exploring the State of the Art in Legal QA Systems	Hybrid Legal QA (Retriever-Reader and Generative Models)	Surveyed modern Legal QA systems, comparing retriever-reader and generative architectures.	Lack of domain-specific datasets and Explainability in existing Legal QA models.	https://arxiv.org/abs/2209.06049
4	2022	A Survey on Legal Judgment Prediction: Datasets, Metrics, Models and Challenges	Deep learning (CNN, BERT, Transformer)	Reviewed datasets, metrics, and ML models for predicting legal judgments and outcomes	Focused on prediction, not full research pipeline (retrieval, summarization, QA).	https://arxiv.org/abs/2204.04859
5	2022	Pre-trained Language Models for the Legal Domain: A Case Study on Indian Law	Legal-BERT fine-tuning)	Fine-tuned pre-trained transformer models on Indian legal corpora for classification and retrieval.	No support for summarization, QA, or Explainability layers.	https://arxiv.org/abs/2209.06049

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3639-3643
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0475

4. Objectives

The objectives of the AI-Driven Legal Research Engine are focused on improving the efficiency, accuracy, and accessibility of legal research in commercial courts. The key objectives include:

- To retrieve relevant legal documents based on contextual meaning rather than simple keyword matching.
- To generate clear and concise summaries of lengthy legal judgments for quick understanding.
- To provide accurate and context-aware answers to complex legal queries.
- To reduce manual effort and research time by automating legal information extraction and analysis.
- To offer a user-friendly platform that improves accessibility and supports better legal decision-making.

5. Societal Contribution Of The System

The AI-Driven Legal Research Engine contributes to society by making legal information more accessible and easier to understand for a wider audience. It reduces the dependency on complex manual research, allowing legal professionals to deliver faster and more accurate outcomes in commercial disputes. By simplifying lengthy legal documents into clear and concise summaries, the system helps individuals better understand their legal rights and obligations. The system also supports improved judicial efficiency by minimizing delays caused by time-consuming research processes. Faster access to relevant case laws and judgments can contribute to quicker resolution of cases, thereby reducing court backlogs[6-10]. Additionally, it promotes fairness and consistency in legal decision-making by providing structured and reliable information. For students and researchers, the platform serves as a valuable learning tool by enabling easy exploration of legal concepts and case studies. It also encourages transparency by making legal knowledge more open and understandable. Overall, the system empowers both legal professionals and the general public, fostering a more informed and efficient legal ecosystem.

6. Results and Discussion

6.1.Results

The evaluation of the AI-driven legal research system was conducted using RAGAS metrics across three query types: simple (Type A), complex multi-domain (Type B), and dynamic (Type C). For Type A queries, both Traditional RAG (TR) and Agentic Multi-Section RAG (AMR) achieved similar performance, with high context precision, recall, and faithfulness (around 0.85–0.89). However, for Type B and Type C queries, TR showed a significant decline in performance (precision ~0.42–0.45, faithfulness ~0.48–0.50). In contrast, AMR maintained strong results across all metrics (precision ~0.78–0.80, faithfulness ~0.86–0.87). A user study with 56 participants further supported these findings. 82% found the system's answers clear. 80% perceived high accuracy. 78% trusted the outputs. 85% reported ease of use.These results demonstrate both strong quantitative performance and positive user acceptance.

6.2. Discussion

The findings indicate that traditional RAG systems are sufficient for simple legal queries but struggle with complex and evolving legal scenarios. The notable drop in performance for Type B and C queries highlights limitations in handling multi-source reasoning and dynamic legal contexts. In contrast, the Agentic RAG architecture significantly improves performance in these challenging cases. Higher context precision and recall suggest more accurate retrieval of relevant legal information, while improved faithfulness reflects better alignment between generated responses and supporting evidence. Additionally, strong user feedback on clarity, accuracy, trust, and usability reinforces the system's practical effectiveness. Overall, the proposed approach enhances the reliability and efficiency of legal research, making it well-suited for commercial court applications.

Conclusion

This paper presented an AI-Driven Legal Research Engine designed to enhance the efficiency and accuracy of legal research in commercial courts. By integrating semantic search, retrieval-augmented generation, and explainable AI techniques, the proposed system overcomes the limitations of traditional keyword-based research methods. The experimental results demonstrate improved relevance, faster information retrieval, and transparent reasoning. The system effectively supports the analysis of large volumes of legal documents and helps legal professionals identify relevant precedents and statutory provisions with greater precision. Moreover, the inclusion of explainable AI strengthens trust and accountability in AI-assisted legal research. In the future, the system can be extended to support additional regional languages, larger datasets, and real-time legal updates, further improving its applicability within the Indian judicial ecosystem.

References

[1]. Magesh, V., Surani, F., Dahl, M., Suzgun, M., Manning, C. D., and Ho, D. E. (2024).Hallucination-free? Assessing the reliability of leading AI legal research tools. arXiv preprint.

[2]. Modi, A., Kalamkar, P., Karn, S., Tiwari, A., Joshi, A., Tanikella, S. K., Guha, S. K., Malhan, S., and Raghavan, V. (2023). SemEval 2023 task 6: LegalEval – understanding legal texts. In Proceedings of the International Workshop on Semantic Evaluation (SemEval 2023).

[3]. Abdallah, A., Piryani, B., and Jatowt, A. (2023). Exploring the state of the art in legal QA systems. arXiv preprint.

[4]. Cui, J., Shen, X., Nie, F., Wang, Z., Wang, J., and Chen, Y. (2022). A survey on legal judgment prediction: datasets, metrics, models and challenges. Artificial Intelligence Review.

[5]. Paul, S., Mandal, A., Goyal, P., and Ghosh, S. (2022). Pre-trained language models for the legal domain: a case study on Indian law. In Proceedings of the Natural Legal Language Processing Workshop.

[6]. Niyogi, M. and Bhattacharya, A. (2024). Paramanu-Ayn: An efficient novel generative and instruction-tuned language model for Indian legal case documents. Gyan AI Research and MBZUAI, in collaboration with IIT Kanpur.

[7]. Lai, J., Gan, W., Wu, J., Qi, Z., and Yu, P. (2023). Large language models in law. ACM Computing Surveys.

[8]. Choi, J. H. (2023). How to use large language models for empirical legal research. Journal of Institutional and Theoretical Economics (forthcoming).

[9]. Hussain, A. and Thomas, A. (2024). Large language models for judicial entity extraction: a comparative study. In Proceedings of the International Conference on Legal Knowledge and Information Systems.

[10]. Zadgaonkar, A. and Agrawal, A. (2021). An overview of information extraction techniques for legal document analysis and processing. International Journal of Electrical and Computer Engineering.

[11]. Erbacher, R. F. and Mulholland, J. (2007). Identification and localization of data types within large-scale file systems. In Second International Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE), pages 55–70. IEEE.

[12]. Khan, S., Zakir, M., Bashir, S., and Ali, R. (2024). Artificial intelligence and machine learning in legal research: a comprehensive analysis. Artificial Intelligence Review