

RAG Based Legal Document Assistant with Verifiable Citations and Multilingual Support

Gayatri Sampat Gavhane¹, Prof. G.R.Sarode², Dr. H.B.Jadhav³

Student, Department of Computer Engineering,¹

Guide, Department of Computer Engineering,²

Co-Guide, Department of Computer Engineering³,

Adsul Technical Campus, Chas, Ahilyanagar, India

Abstract: The increasing availability of legal information in digital form has created a need for intelligent systems that can provide reliable and understandable access to legal documents. Conventional language models may generate fluent responses but can produce unsupported information, particularly when answering complex legal questions or working with multilingual legal content. This paper proposes a Retrieval Augmented Generation based Legal Document Assistant designed to provide grounded responses with verifiable citations. The proposed framework combines multilingual query processing, semantic and lexical document retrieval, relevance based reranking, context aware answer generation, and citation verification within a unified architecture. User queries are first processed according to their language and transformed into representations suitable for multilingual retrieval. Relevant sections from a collection of legal documents are then retrieved and supplied to the language model as supporting context. Each generated response is associated with source references that allow users to trace the information back to the relevant legal document and section. A citation verification mechanism is introduced to identify unsupported statements and improve the reliability of generated answers. The proposed approach is evaluated using retrieval effectiveness, answer relevance, citation correctness, citation coverage, and groundedness across multiple languages. The framework aims to improve accessibility to legal information while reducing unsupported generation and making AI generated legal responses more transparent and traceable. The proposed architecture can serve as a foundation for multilingual legal information systems where accuracy, source verification, and responsible use of artificial intelligence are essential.

Keywords: Retrieval Augmented Generation, Legal Document Assistant, Multilingual Information Retrieval, Large Language Models, Verifiable Citations, Legal Question An

I. INTRODUCTION

The increasing availability of digital legal documents has changed how people access legal information. Traditional retrieval often requires users to examine lengthy acts, regulations, and judgments to locate relevant provisions. A RAG based legal assistant allows users to ask questions in natural language and receive answers based on retrieved legal content. However, language models may generate unsupported information, making source verification important in legal applications.

Recent research has explored RAG and Large Language Models for legal question answering and domain specific information retrieval. The Legal Query RAG study demonstrates specialized retrieval and evaluation mechanisms, while another study combines language models with RAG for legal information. These studies show the potential of RAG, but there remains a need for systems that combine multilingual interaction with direct verification of supporting sources.

The objective of this work is to design a RAG based Legal Document Assistant with Verifiable Citations and Multilingual Support. The proposed system focuses on five functions: (i) multilingual legal query processing, (ii) legal



document retrieval, (iii) context based answer generation, (iv) citation verification, and (v) source grounded response generation. The system aims to provide understandable legal answers while allowing users to trace information to relevant legal documents.

II. OBJECTIVES AND CONTRIBUTIONS

- To develop a multilingual legal query processing mechanism for handling legal questions in different languages.
- To retrieve relevant legal information using lexical and semantic retrieval techniques.
- To generate answers based on retrieved legal documents and supporting evidence.
- To provide verifiable citations that allow users to trace information to its original legal source.
- To verify citation correctness and answer groundedness to reduce unsupported responses.
- To evaluate the proposed framework using retrieval effectiveness, answer relevance, citation correctness, citation coverage, and groundedness.

III. RELATED WORK

The first reference paper considered in this work presents a Legal Query RAG framework that combines retrieval augmented generation with specialized legal models, hybrid retrieval, reranking, and evaluation mechanisms. Its main contribution is the use of retrieval and evaluation components to improve the relevance and reliability of legal question answering [1]. The present work differs by focusing on multilingual legal retrieval and verifiable citations that allow users to trace generated information to its supporting legal source.

The second reference study explores the combination of fine tuned language models and RAG for legal information, particularly in domain specific and multilingual settings. It highlights the usefulness of RAG for accessing changing legal information and working with smaller legal document collections [2]. The proposed system extends this direction by integrating multilingual query processing with citation verification and grounded response generation.

Other research in legal NLP and RAG demonstrates the importance of document retrieval, semantic representations, domain adaptation, and evaluation for legal question answering. These works motivate the need for a system that not only generates relevant answers but also provides transparent evidence for the information presented. Therefore, this paper focuses on combining multilingual retrieval, context based generation, and verifiable citation mechanisms within a unified legal document assistant

IV. RESEARCH GAP

Existing emphasis	Observed limitation	Proposed approach
Legal document retrieval	Retrieval alone may not ensure reliable answers	Retrieval with grounded answer generation
Fine tuned legal language models	Requires domain specific training and resources	RAG based approach using retrieved legal context
Semantic retrieval	May miss exact legal terms or provisions	Combined lexical and semantic retrieval
Legal question answering	Generated answers may contain unsupported information	Citation grounded response generation
RAG based legal systems	Source verification is not always directly provided	Verifiable citations linked to legal sources
Monolingual legal interaction	Limited accessibility for multilingual users	Multilingual legal query processing

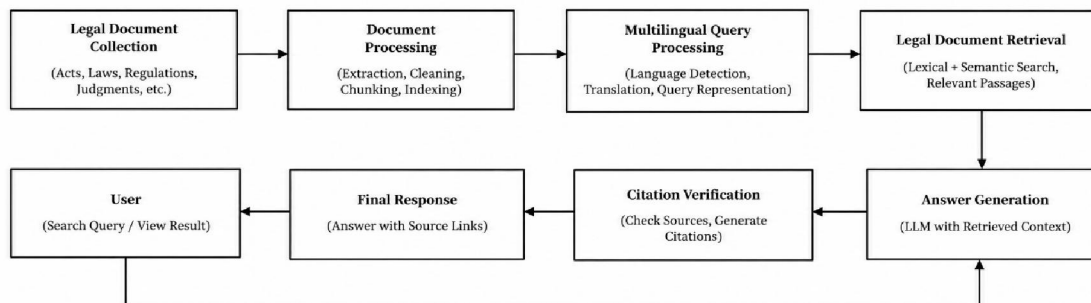


Existing emphasis	Observed limitation	Proposed approach
Retrieved context evaluation	Does not directly verify every generated claim	Citation correctness and groundedness verification
Static legal information	Legal documents and regulations may change	Retrieval from an updated legal document collection

V. PROPOSED SYSTEM

The proposed RAG based Legal Document Assistant is organized into five functional layers: legal document collection, document processing, multilingual query processing, legal document retrieval, and answer generation with citation verification. The document collection layer stores acts, laws, regulations, judgments, and other legal documents. These documents are extracted, cleaned, divided into suitable chunks, and indexed for efficient retrieval. The multilingual query layer accepts legal questions in different languages and prepares them for retrieving relevant legal information. After a user submits a query, the system identifies the language and performs multilingual query processing. The retrieval layer combines lexical and semantic search to identify relevant legal passages. These passages are provided as context to the language model for answer generation. The citation verification layer checks the generated response against the retrieved legal sources and provides verified citations, allowing users to trace the answer back to the supporting legal documents..

• Proposed layered architecture of the RAG based Legal Document Assistant.



VI. HARDWARE COMPONENTS

Component	Function	Design consideration
Computer / Server	Hosts the legal document collection, RAG system, and language model	Sufficient processing capacity
Legal Document Storage	Stores acts, laws, regulations, judgments, and other legal documents	Adequate storage capacity and secure access
Processor	Performs document processing, retrieval, and answer generation	Suitable computational performance
RAM	Supports document processing and language model execution	Sufficient memory for system operations
Network Connection	Provides access to the application and required online resources	Reliable and stable connectivity
Vector Database	Stores document embeddings for semantic retrieval	Efficient similarity search



Component	Function	Design consideration
Database	Stores processed legal documents and metadata	Secure and organized data management
User Device	Allows users to submit legal queries and view responses	Simple and accessible interface
Power Supply	Provides power to the computer or server	Stable and reliable power source

VII. SOFTWARE COMPONENTS

- Legal document collection module for storing acts, laws, regulations, judgments, and other legal documents..
- Document processing module for text extraction, cleaning, chunking, and indexing of legal documents.
- Multilingual query processing module for language detection and processing of legal queries in different languages.
- Retrieval module using lexical and semantic search to identify relevant legal passages.
- RAG based answer generation module for generating answers using the retrieved legal context.
- Citation verification module for checking supporting sources and linking citations to the relevant legal content.
- Web or mobile user interface for submitting legal queries and viewing answers with source citations.

VIII. SYSTEM METHODOLOGY

- **Legal Document Processing and Retrieval**
The system collects legal documents such as acts, laws, regulations, and judgments and processes them through text extraction, cleaning, and chunking. The processed documents are stored for efficient retrieval. When a user submits a legal query, lexical and semantic retrieval techniques are used to identify relevant legal passages from the document collection
- **Multilingual Query Processing**
The system accepts legal queries in multiple languages and identifies the language of the submitted query. The query is processed into a suitable representation for retrieval while preserving the legal meaning of the original question. This allows users to interact with the legal document assistant using their preferred language.
- **Answer Generation and Citation Verification**
The retrieved legal passages are provided as context to the language model for generating the final response. The generated answer is checked against the retrieved content to ensure that the information is supported by relevant legal sources. Verifiable citations are then attached to the response so that users can trace the information back to the corresponding legal document.

IX. PROPOSED ALGORITHM

1. Initialize the legal document collection, processing module, multilingual query interface, retrieval system, language model, and citation verification module..
2. Load and prepare the available legal documents for processing.
3. Extract text from the legal documents and perform cleaning, normalization, and suitable chunking.
4. Generate and store representations of the processed legal document sections for efficient retrieval.
5. Accept a legal query from the user and identify the language of the query.
6. Process the query into a suitable representation while preserving its legal meaning.
7. Retrieve relevant legal passages using lexical and semantic retrieval techniques.
8. Provide the retrieved legal content as context to the language model for answer generation..
9. Verify the generated answer against the retrieved legal content and identify supporting sources.
10. Attach verifiable citations to the relevant information in the generated response.



11. Present the final grounded answer with source references to the user and return to the query processing stage for the next request..

X. SECURITY AND PRIVACY DESIGN

Because a legal document assistant handles sensitive legal information and user queries, security and privacy are important design requirements. The proposed system uses secure access control, protected communication, and controlled handling of legal documents and user information. Legal documents should be stored securely, user access should be authenticated, and unnecessary personal or query information should not be retained.

1. Use encrypted communication between the user interface, application server, retrieval system, and language model.
2. Use authenticated user access and secure credentials to prevent unauthorized system access.
3. Protect stored legal documents, document metadata, and user queries from unauthorized access.
4. Provide controlled access to legal documents and system resources according to user permissions.
5. Verify retrieved sources before displaying citations with generated legal answers.
6. Keep user queries, generated responses, and system logs only for necessary operational and security purposes.
7. Provide mechanisms for updating or removing outdated legal documents to maintain reliable and current information.

XI. EVALUATION METHODOLOGY

A publication quality implementation of the proposed legal document assistant should be evaluated using reproducible experiments rather than unsupported performance claims. The following metrics are recommended. Actual measured values should be inserted after testing the prototype.

Metric	Measurement method	Target interpretation
Retrieval Hit Rate	Measure whether relevant legal passages are retrieved for a set of test queries	Higher is better
Mean Reciprocal Rank	Measure the ranking position of the first relevant legal passage	Higher is better
Answer Relevance	Evaluate whether the generated answer directly addresses the legal query	Higher relevance is better
Groundedness	Compare the generated answer with the retrieved legal context	Higher groundedness is better
Citation Correctness	Check whether each citation correctly supports the corresponding information in the answer	Higher correctness is better
Citation Coverage	Measure the proportion of factual claims that are supported by citations	Higher coverage is better
Multilingual Query Accuracy	Evaluate legal queries submitted in different supported languages	Higher accuracy is better
Response Time	Measure time from query submission to final cited response	Lower response time is better
Availability	Record successful query processing over repeated trials	High service availability

XII. TEST SCENARIOS

- Legal query using the supported primary language.
- Legal query submitted in a different supported language.



- Query requiring retrieval of a specific legal provision.
- Query with multiple relevant legal documents.
- Query containing legal terminology and exact provisions.
- Query where relevant information is available in retrieved documents.

XIII. RESULTS AND DISCUSSION

The proposed architecture is designed to be evaluated as a functional legal document assistant rather than as a purely conceptual system. A complete experimental section should report the measured values from the test scenarios in Section 8. The important evaluation factors include retrieval effectiveness, answer relevance, citation correctness, multilingual query processing and response time. Higher retrieval accuracy and groundedness indicate that the system is able to identify and use relevant legal information. Citation verification should also be evaluated separately to determine whether the generated answers are properly supported by the retrieved legal sources.

For multilingual queries, the system should be evaluated using legal questions in the supported languages while preserving the meaning of the original query. The RAG based architecture can reduce unsupported responses by providing relevant legal passages as context to the language model. However, incorrect retrieval or insufficient source information may affect the quality of the generated answer. Therefore, retrieval performance, answer groundedness and citation coverage should be reported separately rather than represented by a single accuracy value.

XIV. ADVANTAGES AND LIMITATIONS

- **Advantages**
 1. Multilingual access to legal information through natural language queries.
 2. Retrieval of relevant legal information from legal documents.
 3. Context based answer generation can reduce unsupported responses.
 4. Verifiable citations allow users to trace answers to supporting legal sources.
 5. Modular design permits integration of document processing, retrieval, language models, and citation verification
- **Limitations**
 1. Performance depends on the quality and availability of the legal document collection.
 2. Incorrect or incomplete retrieval can affect the quality of generated answers.
 3. Multilingual processing may be less accurate for languages with limited legal language resources.
 4. Legal documents may change over time, requiring regular updates to maintain reliable information.
 5. Language models may still produce unsupported information if retrieved context is insufficient or incorrectly interpreted.

XV. FUTURE SCOPE

1. Integration of larger and continuously updated legal document collections.
2. Support for additional regional and international languages for multilingual legal queries.
3. Development of improved multilingual legal embeddings for better retrieval across languages.
4. Integration of advanced citation verification for claim level source checking.
5. Development of real time legal document updating to reflect changes in laws and regulations.
6. Integration of legal expert feedback for improving answer quality and system reliability.
7. Development of benchmark datasets for evaluating multilingual legal retrieval, groundedness, and citation correctness..



XVI. CONCLUSION

This paper proposed a RAG based Legal Document Assistant that combines multilingual legal query processing, legal document retrieval, context based answer generation, and verifiable citation mechanisms in a unified architecture. The proposed system processes legal documents into suitable sections, retrieves relevant legal information using lexical and semantic retrieval techniques, and provides the retrieved context to a language model for generating grounded responses. Unlike conventional legal question answering systems, the proposed approach focuses on providing citations that allow users to trace generated information to its supporting legal sources. Multilingual query processing improves accessibility by allowing users to interact with the system in different supported languages. Citation verification, groundedness checking, secure access, and controlled handling of legal information are incorporated to improve reliability and responsible use. Future evaluation should measure retrieval effectiveness, answer relevance, citation correctness, citation coverage, multilingual query accuracy, response time, and system availability under realistic legal information retrieval scenarios. The proposed architecture can serve as a foundation for accessible, transparent, and source grounded legal information retrieval applications.

REFERENCES

1. R. S. M. Wahidur, S. Kim, H. Choi, D. S. Bhatti, and H. N. Lee, "Legal Query RAG," IEEE Access, vol. 13, 2025.
2. P. S. García-Montero, P. Vizcaíno, I. G. Reyes-Chacón, and M. E. Morocho-Cayamcela, "Legal AI for All: Reducing Perplexity and Boosting Accuracy in Normative Texts With Fine-Tuned LLMs and RAG," IEEE Access, vol. 13, 2025.
3. N. Pipitone and G. Houir Alami, "LegalBench-RAG: A Benchmark for Retrieval-Augmented Generation in the Legal Domain," 2024.
4. S. A. A. Sha, S. Sivanesan, A. K. Madasamy, and N. Binu, "Rhetorical-Role-Aware Retrieval-Augmented Generation for Legal Question Answering over Indian Supreme Court Judgments," 2026.
5. S. K. Nigam, S. K. Mishra, N. Shallum, K. Ghosh, and A. Bhattacharya, "AILQA: Evaluating AI-driven legal question answering systems for the Indian legal system," Artificial Intelligence and Law, 2026.

