CSE 499A

Section 02

Report

Submitted to

Dr. Mohammad Rashedur Rahman [RRn]

Submitted By

Name	ID
Abdullah Al Raiyan	2212712042
Md. Sifat Haque Zidan	2212768042
Rahil Mehnaz	2121843042

A Multi-Agent RAG system for Legal Information Retrieval in Bangladesh

Abdullah Al Raiyan[1], Md. Sifat Haque Zidan[2], Rahil Mehnaz[3], and Dr. Mohammad Rashedur Rahman[4]

Department of Electrical and Computer Engineering
North South University, Dhaka 1229, Bangladesh
[1]abdullah.raiyan@northsouth.edu
[2]sifat.zidan@northsouth.edu
[3]rahil.mehnaz@northsouth.edu
[4]rashedur.rahman@northsouth.edu

Abstract. Understanding legal issues can be difficult for individuals without legal knowledge or access to professional legal advice. Access to legal information remains a significant challenge in developing countries such as Bangladesh, where a large portion of the population lacks legal knowledge and access to professional legal counsel. We propose a Multi-Agent Retrieval Augmented Generation (RAG) chatbot tailored for Bangladeshi legal information retrieval to address this gap. This application employs two specialized agents: a Clarification Agent and a Request Relevant RAG Agent. The Clarification Agent manages the conversational flow to refine the user's initial query. It uses LLMs to assess clarity and generate follow-up questions. The RAG agent takes the (potentially clarified) query and executes the Retrieval-Augmented Generation process, including document retrieval, grading, optional web search, and final answer generation. This system integrates a carefully curated legal dataset compiled by our team from various official government websites, legal blogs, and scholarly articles. The dataset has been thoughtfully selected to ensure it accurately represents the legal landscape of Bangladesh. This chatbot uses a modern LLM app framework, `LangChain`, and a smart workflow engine, `LangGraph`, to coordinate intelligent agents that work together to answer Bangladeshi legal questions accurately and interactively. Through the structured collaboration between autonomous agents, the chatbot handles complex legal queries by combining document retrieval, relevance checking, clarification prompts and conditional web searches. A conversational `Streamlit` based interface enables accessible, real-time interactions. This work demonstrates a scalable and cost-effective approach to democratizing legal knowledge, particularly in low-resource settings like Bangladesh.

Keywords: Multi-Agent · Retrieval Augmented Generation (RAG) · Legal Chatbot · Bangladesh Law · LangChain · LangGraph · Information Retrieval · Natural Language Processing · Low-Resource Settings.

1 Introduction

Legal services are fundamental for protecting individual rights, ensuring social justice and promoting social equity. However, widespread access to reliable legal assistance services remains a significant challenge, especially in underprivileged and densely populated countries like Bangladesh due to the scarcity of legal professionals, lack of public legal awareness, and high associated costs. Even though Bangladesh has made some progress in offering legal aid to its people, the support system is still not enough to meet the growing needs. Many people, especially in rural or underprivileged areas, cannot get help for their simple legal problems or don't know what their legal rights are. This situation brings up an important question: Can smart and automated systems be used to make legal information more available and easier to understand for everyone in Bangladesh?

In recent years, Large Language Models (LLMs) have shown remarkable success across domains such as healthcare, finance, and law. Tools like ChatGPT, LLaMA and other models have been increasingly explored for their potential in the legal domains. However, despite their potential, traditional LLMs frequently generate hallucinated outputs and answers that may seem accurate but are actually incorrect or they are based on outdated information. This can be dangerous in the legal context, where accuracy and up-to-date knowledge are critically important.

To address these limitations, we developed a Multi-Agent Retrieval-Augmented Generation (RAG) chatbot, tailored for Bangladeshi legal information retrieval. Our solution integrates curated datasets including Supreme Court judgments [1], official legislative documents [2], and legal analysis articles [3] into a knowledge base designed to support factual retrieval. By utilizing `LangChain` [5] and `LangGraph` [6] frameworks, the system introduces a multi-agent architecture where the Retriever Agent identifies relevant legal content, evaluates and refines results based on their relevance to the query. Furthermore, the system is equipped with a conditional web search mechanism to extend support when local resources are insufficient.

For user interaction, a `Streamlit`-based conversational interface [7] was developed, allowing individuals to input queries and receive clear, relevant legal information in real time. Through this work, we aim to present a scalable, low-cost blueprint for enhancing public access to legal knowledge in low-resource settings like Bangladesh, potentially transforming how citizens engage with legal information.

2 Related Works

The integration of artificial intelligence and natural language processing technologies into the legal domain has gained significant momentum in recent years with various approaches such as improving the accessibility, reliability, and efficiency of legal services. The development of legal chatbots and AI-driven legal assistance systems has been a growing research focus in recent years. Various approaches have been proposed to address the challenges of legal knowledge retrieval, accurate response generation, and user trust. In this section, we will review some related and most relevant works in this domain:

2.1 ChatLaw: A Multi-Agent Collaborative Legal Assistant

Jiaxi Cui et al. [8] introduced a Mixture-of-Experts (MoE) architecture combined with a multi-agent collaboration framework to address hallucination and accuracy issues in legal question-answering systems in the Chinese legal domain. By integrating a retrieval-augmented generation (RAG) model and specialized expert networks based on legal sub-domains, `ChatLaw` outperformed GPT-4 across multiple benchmarks, emphasizing the role of specialization and grounded retrieval in improving legal LLM performance.

2.2 Building a Legal Dialogue System

Mudita Sharma et al. [9] proposed an architecture leveraging AWS Lex and AWS Lambda to build a legal chatbot capable of assisting users in navigating legal services. Due to the scarcity of public legal dialogue datasets, they collected data via crowdsourcing and user interactions. Their hierarchical bot design first gathered essential user information before providing appropriate legal guidance, demonstrating practical strategies for data collection and bot structuring in low-resource domains.

2.3 The Use of Chatbots in Providing Free Legal Guidance

Ogunsan Isaac et al. [10] critically examined the ethical, legal, and regulatory challenges associated with legal chatbots like DoNotPay. They highlighted the opportunities for increasing access to justice but warned about the risks of misinformation, liability issues, and the potential for unethical practices. The study called for stronger oversight, transparency, and user protection mechanisms when deploying AI in legal advisory contexts.

2.4 Technology Acceptance Model for Lawyer Robots with AI

Ni Xu et al. [11] extended the classic Technology Acceptance Model to understand the factors influencing user adoption of legal chatbots. The study found that perceived ease of use and usefulness were stronger predictors of trust and acceptance than mere regulatory compliance. It underscores the importance of focusing on user experience and credibility when designing legal conversational agents.

2.5 Improving Access to Justice with Legal Chatbots

Marc Queudot et al. [12] developed two retrieval-based chatbots: one aimed at helping immigrants navigate legal procedures, and another for corporate employees seeking legal compliance information. Their work demonstrated that conversational agents could significantly improve accessibility to specialized legal knowledge, particularly for vulnerable or underserved populations.

2.6 A Chatbot for Specialized Domain

Egidia Cirillo et al. [13] tackled the problem of unstructured legal text by proposing a semi-automated pipeline using generative AI. Their approach involved parsing, indexing, and validating large legal corpora to make them usable for retrieval-augmented generation models. This structuring method is critical for improving the accuracy, relevance, and auditability of legal information retrieval systems.

2.7 Transforming legal text interactions

Mohammed Maree et al. [14] presented the development of a legal chatbot trained on a newly created dataset covering Palestinian cooperative law. The system achieved 82% accuracy in answering legal queries and was validated against human expert judgments. This work illustrates how fine-tuning LLMs with domain-specific and localized datasets can yield effective legal assistance tools in niche legal sectors.

2.8 LAWBOT: A Smart User Indian Legal Chatbot

Nikita et al. [15] introduced `LawBot`, a legal chatbot designed to simplify access to Indian legal resources. Combining machine learning and IR techniques, `LawBot` assists users by providing relevant legal advice and references. The project highlighted how technology can help bridge the gap between legal services and common citizens in complex legal environments.

2.9 An Intelligent Conversational Agent for the Legal Domain

Flora Amato et al. [16] introduced `CREA2`, an intelligent agent designed to help users navigate legal concepts, draft legal documents, and propose solutions in disputes such as divorces, inheritances, and corporate divisions. The agent leverages natural language processing (NLP) and semantic search, utilizing SBERT for evaluating query relevance in an unsupervised manner. `CREA2` demonstrates how conversational agents can reduce legal professionals' workloads and improve user access to legal support across digital platforms.

2.10 Legal Solutions - Intelligent Chatbot Using Machine Learning K. R. A et al. [17]

proposed an AI-powered legal chatbot that empowers users with fundamental legal knowledge, real-time attorney consultations, and personalized legal instructions based on location and financial status. By integrating advanced machine learning and NLP techniques, the chatbot autonomously retrieves and processes legal information, aiming to democratize legal resource access and provide rapid, context-specific support to a wide range of users.

2.11 Development of a legal Document AI-Chatbot

Devraj et al. [21] guides beginners in basic legal Chabot development by explaining the processes from the start, gives an idea about `LangChain`, introduces the technique Cosine Similarity, the design of their bot, frontend backend giving us the idea of what our bot might look like and how `LangChain` is connected, and how the texts are broken into small chunks are explained easily.

2.12 RAGged Edges: The Double-Edged Sword of Retrieval-Augmented Chatbots

Feldman et al. [23] explores how Retrieval-Augmented Generation (RAG) can counter hallucinations by integrating external knowledge with prompts. Their results show that RAG increases accuracy in some cases but can still be misled when prompts directly contradict the model's pre-trained understanding. In this paper, they tested the effectiveness of context prompting as used in RAG to determine its effectiveness when compared to the same prompt without context, which we applied in our agentic RAG. Their research has shown that in the presence of complex or misleading search results, a RAG system may often get things wrong. A missing section may lead to hallucinations; an unconventional placement of dates may result in the time from one event being attributed to another.

2.13 RAGAS: Automated Evaluation of Retrieval Augmented Generation

For evaluation, es et al. [24] has shown that the predictions from RAGAs are closely aligned with human predictions, especially for faithfulness and answer relevance.

While the studies mentioned above have made significant contributions in the field of legal chatbots, our work introduces a few unique aspects that set it apart. First and foremost, this work focuses on Bangladeshi legal information, which is often overlooked in many global chatbot systems. Unlike the systems discussed earlier, which generally provide legal guidance in a broad sense, this chatbot is specifically designed for Bangladesh's legal context. Another important difference is the use of a multi-agent system, which allows the chatbot to handle more complex, multi-turn conversations with users. It also goes beyond basic legal information by incorporating features like conditional web searches and relevance checking, which ensure that users get responses based on the most up-to-date legal data. Additionally, this chatbot includes clarification prompts, a feature that helps users refine their questions, improving the quality of the legal advice given. These tailored features make this work stand out as a more specialized and region-specific solution, particularly suited to the needs of Bangladeshi users, which differentiates it from the more generalized systems found in previous research.

3 Methodology

In this project, we developed a Multi-Agent Retrieval-Augmented Generation (RAG) chatbot system specialized for Bangladeshi legal information retrieval. The overall methodology can be divided into several key stages:

3.1 Data Collection and Preparation

At first, we collected a set of legal documents, mainly PDF files, from sources like Bangladesh Supreme Court judgments [1], Laws of Bangladesh [2], and BD Law Post articles [3], BD Law Help [20], etc. These PDFs were loaded using the `PyPDFDirectoryLoader` from the `langchain.document_loaders` module [5]. Additionally, to extract more accurate text from complicated PDFs, we used the `fitz` library from `PyMuPDF` [4].

After loading the documents, we needed to split the long texts into manageable chunks for better retrieval. Here, we used two techniques:

– `RecursiveCharacterTextSplitter` [5] to split text based on certain characters like "\n\n" or ".".
– `SemanticChunker` [5] from LangChain, which uses an embedding model to chunk text more meaningfully based on semantic closeness.

Fig. 1: Data Preparation Workflow.

3.2 Embedding Generation

For generating dense vector representations of each text chunk, we used OpenAI Embeddings [18]. This helped in converting text into high-dimensional vectors, making it easier to perform similarity search later on.

3.3 Vector Store Creation

After getting embeddings, we stored them in a vector database called `Chroma` [19]. `Chroma` is a simple and lightweight vector store that supports similarity search. We set up the `ChromaDB` locally using the `LangChain Chroma` wrapper. This made it easy to retrieve the most relevant chunks when a user asks a query.

3.4 Multi-Agent System with LangGraph

One of the core innovations of this project was using `LangGraph` [6] to build a graph-based, multi-agent system. Unlike traditional RAG systems with a single retrieval loop, we designed two specialized agents:

- **Clarification Agent:** This agent orchestrates the initial user interaction phase, focusing on query refinement. It employs specialized Large Language Model (LLM) configurations to perform structured assessments of query clarity (*assessment_llm*) and, when necessary, generates targeted follow-up questions (*question_gen_llm*) to resolve ambiguity or elicit sufficient detail prior to information retrieval.

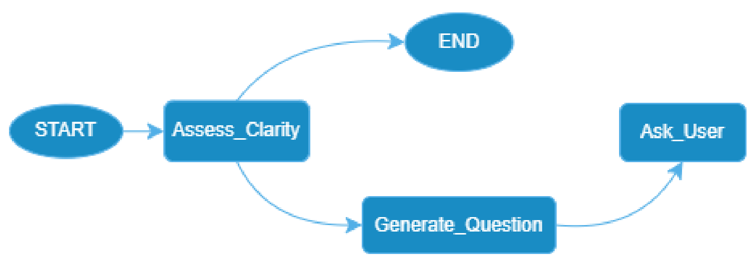

Fig. 2: Clarification Agent Workflow.

- **RAG Agent:** Responsible for executing the core Retrieval-Augmented Generation process upon receiving a sufficiently clear query. This agent integrates several components:
  - It utilizes a *Vector Database Retriever* (`vector_db.as_retriever`) for primary information retrieval from a specialized corpus (Chroma DB).
  - It employs an LLM (*rewriter_llm*) to potentially rephrase the query for optimal retrieval performance against both the vector database and external search engines.
  - It leverages a structured LLM (*structured_llm_grader*) to evaluate the relevance of retrieved document chunks.
  - It incorporates an external web search capability via the *Tavily Search Tool* (`TavilySearchResults`) as a fallback or supplementary information source.

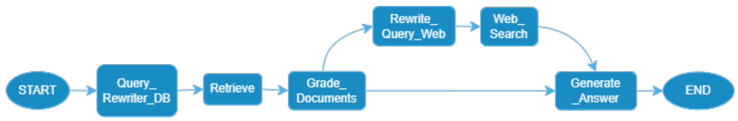

Fig. 3: RAG Agent Workflow.

- Finally, it synthesizes a comprehensive answer using a generative LLM (*llm*), grounded in the context provided by the retrieved, graded, and potentially web-augmented information.

In aggregate, this multi-agent workflow comprises a total of **two distinct agents**. These agents collectively leverage **seven key functional components**: two external tools (one vector database retriever and one web search tool) and five specialized LLM configurations, each assigned a specific role within the clarification, retrieval, grading, or generation stages of the process. Both agents were implemented using `LangGraph`'s `StateGraph`, with `add_node`, `add_edge`, and `add_conditional_edges` functions [6].

3.5 Query Rewriting

Before sending the user's raw query to the retriever, a Question Rewriter Node reformulates the query to make it clearer and easier to retrieve relevant information.

3.6 Conditional Web Search

Another important part of the system is the structure for conditional web search. If the local database provides enough information, the system answers directly. And if not, it can be extended to perform a real-time web search for additional information. This decision-making logic was handled using a `should_search` function and conditional routing inside the `LangGraph` [6].

3.7 Language Model and Tool Handling

For language understanding and text generation, we used OpenAI's GPT 4o mini through the `LangChain`'s `ChatOpenAI` wrapper [18]. Several "tools" were defined, such as: retrieval tool (fetch chunks), search tool (for web search), answer tool (format the answer). Agents dynamically call these tools based on the need detected from the conversation flow.

3.8 User Interface

To make the system accessible, we built a prototype `Streamlit` UI [7]. Through this interface, users can input their legal queries and receive relevant answers retrieved and generated through the multi-agent RAG system.

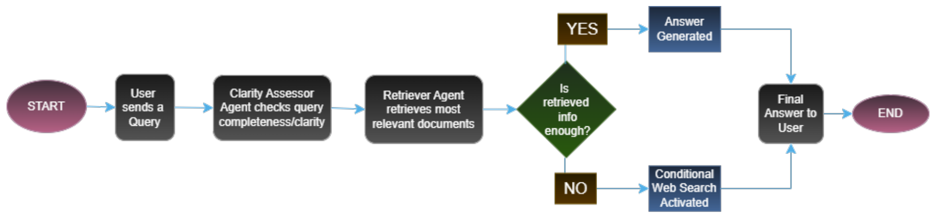

Fig. 4: Detailed Agent Workflow.

4 Results

The developed multi-agent RAG system resulted in a functional chatbot capable of addressing legal queries specific to the Bangladeshi context. An example interaction showcasing the system's ability to handle a user query regarding the Right to Information (RTI) Act is presented in Figure 5.

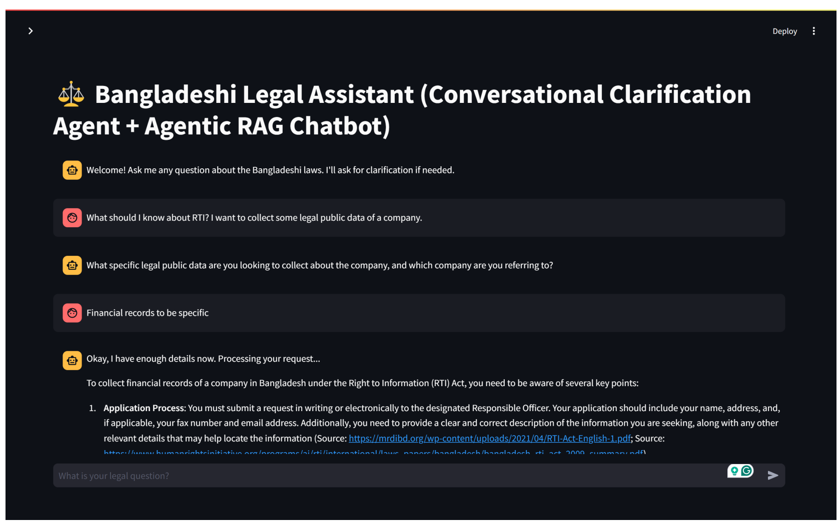

Fig. 5: Example interaction demonstrating a query ("What should I know about RTI? I want to collect some legal public data of a company.") and the corresponding detailed response generated by the multi-agent system.

To quantitatively assess the performance and reliability of the system, we employed the Deepeval framework [26] for evaluation. A curated test suite comprising 10 distinct legal questions relevant to the laws and legal landscape of Bangladesh was used. The evaluation focused on a set of standard RAG metrics designed to measure different facets of the system's retrieval and generation quality. The specific metrics evaluated include:

- **Answer Relevancy:** Measures how pertinent the generated answer is to the user's query. It is calculated as the ratio of relevant statements to the total statements in the answer:

$$\text{Answer Relevancy} = \frac{\text{Number of Relevant Statements}}{\text{Total Number of Statements}}$$

- **Faithfulness:** Assesses whether the generated answer is factually consistent with the retrieved context documents. It is calculated based on the claims made in the answer versus the provided context:

$$\text{Faithfulness} = \frac{\text{Number of Truthful Claims}}{\text{Total Number of Claims}}$$

- **Contextual Precision:** Evaluates the quality of the retrieval component by measuring the proportion of relevant documents among the retrieved set, weighted by their rank:

$$\text{Contextual Precision} = \frac{1}{\text{Number of Relevant Nodes}} \sum_{k=1}^{n} \left(\frac{\text{Number of Relevant Nodes Up to Position } k}{k} \times r_k \right)$$

  where r_k indicates if the node at rank k is relevant.
- **Contextual Recall:** Measures the extent to which the retrieved context contains all the necessary information required to formulate the ideal answer (often compared against a ground truth or expected output):

$$\text{Contextual Recall} = \frac{\text{Number of Attributable Statements}}{\text{Total Number of Statements}}$$

  (Note: Deepeval calculates this based on the expected output).
- **Contextual Relevancy:** Assesses the overall relevance of the retrieved context passages to the input query, calculated as:

$$\text{Contextual Relevancy} = \frac{\text{Number of Relevant Statements}}{\text{Total Number of Statements}}$$

  (Note: Deepeval calculates this based on the input query and the retrieved context).

Deepeval utilizes an LLM-as-a-judge approach for these metrics, often providing reasoning alongside the numerical scores.

The average scores obtained across the 10 test questions for these metrics, plus an additional 'Correctness (GEval)' metric reported by the tool, are presented in Figure 6 (based on the provided image data).

The average performance metrics were recorded as follows:

- Correctness (GEval): 0.77

– Answer Relevancy: 1.00
– Faithfulness: 1.00
– Contextual Precision: 0.89
– Contextual Recall: 1.00
– Contextual Relevancy: 0.75

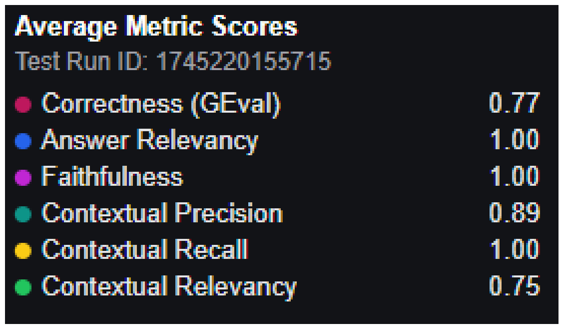

Fig. 6: Summary of average RAG metric scores obtained from the Deepeval evaluation across 10 test questions (Test Run ID: 1745220155715).

These results demonstrate strong performance in several key areas. Perfect scores (1.00) for **Answer Relevancy** and **Faithfulness** indicate that the system consistently generates answers that directly address the user's query and are factually grounded in the retrieved Bangladeshi legal documents. A perfect score in **Contextual Recall** (1.00) suggests that the retriever successfully identified context containing the necessary information to answer the questions comprehensively. **Contextual Precision** (0.89) is also high, showing that the most relevant documents were generally ranked highly within the retrieved set. The lower scores for **Contextual Relevancy** (0.75) and **Correctness (GEval)** (0.77) suggest that while the core necessary information was retrieved and used faithfully, the retrieved context might sometimes contain less relevant segments, or there might be nuances in overall correctness (potentially related to completeness or specific phrasing) not fully captured by the faithfulness metric alone. Overall, the evaluation indicates a robust system, particularly strong in relevance and factual grounding based on the provided context, with potential for refinement in optimizing the conciseness of retrieved context and addressing finer points of correctness.

5 Discussion

The development of this multi-agent RAG system represents a significant step towards enhancing legal information accessibility in Bangladesh, a context characterized by limited resources and low legal literacy among the general populace.

Our findings demonstrate the viability of employing a sophisticated multi-agent architecture, specifically using `LangGraph`, to address the inherent complexities of legal query processing.

The core strength of our approach lies in the division of labor between the Clarification Agent and the RAG Agent. Unlike monolithic RAG systems, the dedicated Clarification Agent proactively manages ambiguity. By assessing query clarity (*assessment_llm*) and prompting users for refinement (*question_gen_llm*), the system aims to improve the quality of the input for the retrieval stage, reducing the likelihood of irrelevant outputs. This initial interaction phase is crucial in the legal domain, where query precision directly impacts the relevance and correctness of the provided information, contributing indirectly to the high performance observed in downstream metrics.

The RAG Agent itself incorporates several layers designed to improve robustness, reflected in the strong evaluation scores. The perfect scores (1.00) achieved for **Answer Relevancy** and **Faithfulness** validate the effectiveness of grounding the generation process (*llm*) in retrieved, domain-specific context obtained via the *Vector Database Retriever* and filtered by the *structured_llm_grader*. This indicates the system successfully avoids hallucination and directly addresses user queries based on the provided legal documents. Furthermore, the perfect **Contextual Recall** (1.00) coupled with high **Contextual Precision** (0.89) suggests that the retrieval mechanism, likely aided by query rewriting (*rewriter_llm*), is effective at identifying and prioritizing documents containing the necessary information. However, the slightly lower scores for **Contextual Relevancy** (0.75) and **Correctness (GEval)** (0.77) highlight areas for improvement. These scores suggest that while the essential information is retrieved and used faithfully, the retrieved context might sometimes include extraneous segments, or the final answer might miss certain nuances affecting overall correctness, reinforcing points made by Feldman et al. [23] about the remaining challenges even in RAG systems. The document grading step aims to mitigate this, but further optimization might be beneficial. The conditional web search capability using the *Tavily Search Tool*, while not explicitly tested in this evaluation set, provides a necessary fallback mechanism to enhance comprehensiveness.

Despite the promising results, several limitations warrant discussion. Firstly, the system's effectiveness remains heavily dependent on the quality and scope of the curated dataset [1–3]. Maintaining comprehensiveness and currency is an ongoing challenge. Secondly, evaluating legal correctness rigorously requires expert human oversight beyond automated metrics like RAGAS [24], which is resource-intensive. Our current reliance on Deepeval [26] provides valuable quantitative insights, but deeper qualitative assessment is needed. Thirdly, dependencies on external LLMs (OpenAI models [18]) introduce considerations of cost, availability, potential biases, and data privacy. Finally, deploying and maintaining such a system affordably at scale in a low-resource setting like Bangladesh requires careful planning.

Nonetheless, the multi-agent framework presented offers a flexible and powerful paradigm. By breaking down the complex task into distinct steps handled by specialized agents and tools, the system achieves a level of sophistication and adaptability, providing a valuable blueprint for leveraging advanced AI to address critical information access gaps in specialized domains like Bangladeshi law.

6 Future Plan for 499B

Building upon the promising initial evaluation results presented in Section 4, particularly the perfect scores for Faithfulness and Answer Relevancy, our future work for the subsequent CSE 499B course will focus on further refinement and validation of the multi-agent system. Currently, the evaluation metric is quite naive and incomplete. A primary goal is to enhance the evaluation methodology [25,26]. While Deepeval provided valuable initial metrics, we plan to incorporate `RAGAS` [24] to gain deeper insights into different facets of the RAG pipeline, such as the quality of generation versus retrieval. Crucially, we aim to conduct expert human evaluations to assess the practical legal correctness and nuance of the system's responses.

Data expansion remains a priority. We will continue curating and integrating more diverse and up-to-date Bangladeshi legal documents into the `Chroma` vector store [19] to improve coverage and potentially boost contextual recall and relevancy further.

We will also explore architectural enhancements within the `LangGraph` framework [6]. Based on the system's current capabilities, adding an agent specialized in identifying and classifying legally similar documents or cases [22], leveraging the existing retrieval mechanisms, could provide significant value. Additionally, exploring the integration of persistent memory for the agents, potentially using an external vector store as suggested by Weng [27], could enable more sophisticated, multi-turn conversational interactions and retain user context across sessions. While incorporating lawyer recommendations was initially considered, we will prioritize core system improvement and validation within the scope of 499B, focusing on enhancing the reliability and scope of the legal information provided.

If possible, we will try to add another agent which suggests the names and contact information of the lawyers related to the experienced in the field of the concern or an agent who identifies classifies similar legal documents or cases or how similar these are [22]. We would consider an agent who has memory can retain recall information by leveraging an external vector store [27].

7 Conclusion

Access to clear and accurate legal information remains a significant barrier in Bangladesh. This paper addressed this challenge by designing, implementing, and evaluating a novel Multi-Agent Retrieval-Augmented Generation (RAG) chatbot specifically tailored for the Bangladeshi legal domain. Leveraging the

LangGraph framework, we developed a distinct two-agent architecture comprising a Clarification Agent for query refinement and a RAG Agent for robust information retrieval, grading, and synthesis.

Our methodology involved curating a specialized legal corpus, employing advanced text splitting and embedding techniques, and orchestrating the agent interactions through a stateful graph. The quantitative evaluation demonstrated the system's effectiveness, achieving perfect scores in Answer Relevancy and Faithfulness, indicating its ability to generate pertinent and factually grounded responses based on the retrieved context. High scores in Contextual Recall and Precision further validated the retrieval component's performance.

The key contribution of this work lies in demonstrating the practical application of a multi-agent RAG system within a low-resource, specialized domain. The structured collaboration between agents allows for handling query ambiguity and improving the reliability of generated answers compared to simpler RAG approaches. While challenges related to dataset maintenance, comprehensive evaluation, and deployment costs remain, this project provides a valuable blueprint for democratizing legal knowledge in Bangladesh. It showcases a scalable and adaptable approach using modern AI frameworks like LangChain and LangGraph to bridge the critical information gap, offering a promising direction for future legal technology development in similar contexts.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

Data Availability

The primary datasets used in this study consist of publicly available legal documents. Supreme Court judgments are available from the Supreme Court of Bangladesh website [1], Laws of Bangladesh are available from the official legislative portal [2], and legal articles were sourced from BD Law Post [3]. Specific datasets derived or compiled during the research can be made available upon reasonable request to the corresponding author, subject to any applicable licensing or privacy constraints.

References

1. Supreme Court of Bangladesh. Supreme Court of Bangladesh. https://www.supremecourt.gov.bd/web/indexn.php
2. Laws of Bangladesh. Laws of Bangladesh. http://bdlaws.minlaw.gov.bd/
3. BD Law Post. BD Law Post Article. https://www.bdlawpost.com/search/label/Articles?max-results=10
4. PyMuPDF. PyMuPDF Documentation. https://pymupdf.readthedocs.io/en/latest/
5. LangChain. LangChain Documentation. https://python.langchain.com/docs/how_to/#document-loaders

6. LangGraph. LangGraph Documentation. https://langchain-ai.github.io/langgraph/
7. Streamlit. Streamlit Documentation. https://docs.streamlit.io/
8. Cui, J., Jiaxi Cui, M., Liu, Z., Chen, B., Yuan, Y., Lin, H., Liu, B., Yao, Y., Tu, T., Yang, L. (2024). Chatlaw: A Multi-Agent Collaborative Legal Assistant with Knowledge Graph Enhanced Mixture-of-Experts Large Language Model. arXiv preprint arXiv:2401.02914. https://arxiv.org/abs/2401.02914
9. Sharma, M., Trivedi-Raiyani, L., Badlani, A., Muthusamy, V. (2021). Building a Legal Dialogue System: Development Process, Challenges and Opportunities. arXiv preprint arXiv:2109.12654. https://arxiv.org/abs/2109.12654
10. Isaac, O. J., Nwabueze, J. (2025). The Use of Chatbots in Providing Free Legal Guidance: Benefits and Limitations. ResearchGate. [Provide more specific publication details if available, e.g., journal, conference]
11. Xu, N., Wang, K.-J., Li, C.-Y. (2022). Technology Acceptance Model for Lawyer Robots with AI: A Quantitative Survey. In: *HCI International 2022 – Late Breaking Work. HCII 2022. Lecture Notes in Computer Science*, vol 13526. Springer, Cham. https://doi.org/10.1007/978-3-031-22107-7_27
12. Queudot, M., Canton, É., Masse, J.-J. (2020). Improving Access to Justice with Legal Chatbots. *Information*, **11**(5), 276. https://doi.org/10.3390/info11050276
13. Cirillo, E., Fersini, E., Gatta, M., Muscio, A. (2024). A Chatbot for Specialized Domain. In: *Intelligent Systems and Applications. IntelliSys 2023. Lecture Notes in Networks and Systems*, vol 851. Springer, Cham. https://doi.org/10.1007/978-3-031-47721-7_12
14. Maree, M., Abu-Qauod, R., Tuffaha, B. (2023). Transforming legal text interactions: leveraging natural language processing and large language models for legal support in Palestinian cooperatives. *Artificial Intelligence and Law*. https://doi.org/10.1007/s10506-023-09358-0
15. Nikita, Srivastav, E., Patel, A., Singh, A., Sharma, R., Rana, D. P. (2024). LAWBOT: A Smart User Indian Legal Chatbot using Machine Learning Framework. In *2024 3rd International Conference on Innovative Sustainable Computational Technologies (CISCT)* (pp. 1-6). IEEE. https://doi.org/10.1109/CISCT61344.2024.10554004
16. Amato, F., Fersini, E., Gatta, M., Sciarrone, C. (2023). An Intelligent Conversational Agent for the Legal Domain. In *Proceedings of the 15th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management - KDIR* (pp. 236-243). SCITEPRESS – Science and Technology Publications. https://doi.org/10.5220/0012194000003598
17. A, K. R., S, K., R, R. (2023). Legal Solutions - Intelligent Chatbot using Machine Learning. In *2023 International Conference on Disruptive Technologies (ICDT)* (pp. 374-378). IEEE. https://doi.org/10.1109/ICDT57929.2023.10150955
18. OpenAI. OpenAI Platform: Embeddings guide. https://platform.openai.com/docs/guides/embeddings
19. Chroma. Chroma Documentation. https://docs.trychroma.com/docs/overview/introduction
20. LAW HELP BD. Penal Codes section. https://lawhelpbd.com/category/cpc/
21. Devaraj, P. N., P V, R. T., Gangrade, A., R, M. K. (2023). Legal Chatbot using LangChain. arXiv preprint arXiv:2311.12719. https://arxiv.org/abs/2311.12719
22. Siino, M., Falco, M., Croce, D., Rosso, P. (2025). Large Language Models for the Legal Domain: A Survey. *IEEE Access*, **13**, 19533-19554. https://doi.org/10.1109/ACCESS.2025.3533217

23. Feldman, P., Foulds, J. R., Pan, S. (2024). RAGged Edges: The Double-Edged Sword of Retrieval-Augmented Chatbots. arXiv preprint arXiv:2403.01193. https://arxiv.org/abs/2403.01193
24. Es, S., James, J., Espinosa-Anke, L., Schockaert, S. (2023). RAGAS: Automated Evaluation of Retrieval Augmented Generation. arXiv preprint arXiv:2309.15217. https://arxiv.org/abs/2309.15217
25. Ragas Documentation. Evaluation Sample Concepts. https://docs.ragas.io/en/stable/concepts/components/eval_sample/
26. DeepEval Documentation. Evaluation Introduction. https://www.deepeval.com/docs/evaluation-introduction
27. Weng, L. (2023). LLM-powered Autonomous Agents. Lil'Log Blog Post. https://lilianweng.github.io/posts/2023-06-23-agent/