A Retrieval-Augmented Chatbot for Legal Question Answering

Mai Đức Huy
Thai Nguyen University of Information and Communication Technology
Thai Nguyen University, Thai Nguyen, Vietnam
student547@ictu.edu.vn

***Abstract*—This undergraduate thesis investigates a legal question-answering chatbot using retrieval-augmented generation. Legal questions require access to relevant documents and careful handling of generated answers. The author studies retrieval-augmented generation, prepares legal text for search, builds a question-answering flow, and implements a Streamlit interface. The thesis documents data preparation, retrieval, and response generation. The report presents a working chatbot demonstration for legal questions. Its conclusion says the project established the main retrieval and response pipeline. Query rewriting, reranking, multiple candidate answers, user accounts, conversation memory, and answer ratings are listed as future developments. The manuscript does not establish legal correctness across an independently labeled question set or suitability for professional legal advice. Answers may be incomplete or wrong when retrieval fails or source documents are outdated. The contribution is a documented student study or prototype whose stated scope can guide further testing.**

***Index Terms*—legal chatbot, retrieval-augmented generation, Streamlit, question answering**

I. Introduction

Legal questions require access to relevant documents and careful handling of generated answers. This thesis investigates a legal question-answering chatbot using retrieval-augmented generation. It begins from a practical engineering need and explains the technologies and workflow chosen for the case. The objective is to document a student implementation or comparative investigation, together with the evidence available in the manuscript. Broader operational or economic effects require separate measurements and are not assumed from the existence of a prototype [1]. Related published studies provide context for the subject and methods [2], [3], [4], [5], [6], [7], [8], [9], [10]. These contextual works were selected for this English summary; they are not asserted to appear in the original thesis bibliography.

II. Methods

The author studies retrieval-augmented generation, prepares legal text for search, builds a question-answering flow, and implements a Streamlit interface. The thesis documents data preparation, retrieval, and response generation. The source is summarized through its stated design, implementation material, examples, and conclusion. Features shown in the manuscript are distinguished from proposals for later development. This review does not add a new experiment, fill missing measurements with estimates, or treat a planned capability as an observed result. The source thesis remains the provenance for the reported work [1].

Supervisor: Trần Văn Khánh, Thai Nguyen University of Information and Communication Technology (tvkhanh@ictu.edu.vn).

III. Results

The report presents a working chatbot demonstration for legal questions. Its conclusion says the project established the main retrieval and response pipeline. Query rewriting, reranking, multiple candidate answers, user accounts, conversation memory, and answer ratings are listed as future developments. The report's screens, configurations, examples, or measurements constitute the evidence for this summary. A capability described only as a design intention is not counted as a validated operational outcome. No additional performance, adoption, financial, or safety figure has been introduced here. The result is interpreted within the scope of the student's documented setup [1].

IV. Discussion

The manuscript does not establish legal correctness across an independently labeled question set or suitability for professional legal advice. Answers may be incomplete or wrong when retrieval fails or source documents are outdated. Further evaluation could test citation grounding, answer accuracy, document currency, refusal behavior, and feedback from qualified legal reviewers. These limitations affect how the evidence should be read. The current manuscript can support understanding of the design and reported demonstration, while independent tests would be needed for stronger claims. Any later deployment should document its environment and separate observed outcomes from intended benefits [1].

V. Conclusion

The thesis documents a legal question-answering chatbot using retrieval-augmented generation. The report presents a working chatbot demonstration for legal questions. Its conclusion says the project established the main retrieval and response pipeline. Query rewriting, reranking, multiple candidate answers, user accounts, conversation memory, and answer ratings are listed as future developments. The manuscript does not establish legal correctness across an independently labeled question set or suitability for professional legal advice.

Answers may be incomplete or wrong when retrieval fails or source documents are outdated. Further evaluation could test citation grounding, answer accuracy, document currency, refusal behavior, and feedback from qualified legal reviewers. Accordingly, the present contribution is a source-grounded account of the student work and its tested or illustrated scope. Claims about broader effectiveness should be assessed in later work using explicit evaluation conditions [1].

References

[1] Mai Đức Huy, ”A Retrieval-Augmented Chatbot for Legal Question Answering,” internal undergraduate thesis (original in Vietnamese), Thai Nguyen University of Information and Communication Technology, 2024.

[2] Jorge Martinez-Gil, ”A survey on legal question–answering systems,” Computer Science Review, 2023. doi:10.1016/j.cosrev.2023.100552.

[3] Biralatei Fawei et al., ”A Semi-automated Ontology Construction for Legal Question Answering,” New Generation Computing, 2019. doi:10.1007/s00354-019-00070-2.

[4] Haoxi Zhong et al., ”JEC-QA: A Legal-Domain Question Answering Dataset,” Proceedings of the AAAI Conference on Artificial Intelligence, 2020. doi:10.1609/aaai.v34i05.6519.

[5] Oleksandr Kolomiyets et al., ”A survey on question answering technology from an information retrieval perspective,” Information Sciences, 2011. doi:10.1016/j.ins.2011.07.047.

[6] Matthew W. Bilotti, ”Linguistic and semantic passage retrieval strategies for question answering,” ACM SIGIR Forum, 2011. doi:10.1145/1924475.1924495.

[7] Yongqi Li et al., ”Generative retrieval for conversational question answering,” Information Processing & Management, 2023. doi:10.1016/j.ipm.2023.103475.

[8] Seok-Hoon Kang et al., ”M-RAG: Enhancing Open-domain Question Answering with Metadata Retrieval-Augmented Generation,” Journal of the Korea Institute of Information and Communication Engineering, 2023. doi:10.6109/jkiice.2023.27.12.1489.

[9] Shamane Siriwardhana et al., ”Improving the Domain Adaptation of Retrieval Augmented Generation (RAG) Models for Open Domain Question Answering,” Transactions of the Association for Computational Linguistics, 2023. doi:10.1162/tacl_a_00530.

[10] Ziyi Fan et al., ”Question & answering system based on retrieval,” Applied and Computational Engineering, 2023. doi:10.54254/2755-2721/4/20230477.