AttorneyGPT – Multilingual Generative Artificial Intelligence Law Chatbot using Retrieval Augmented Generation

A Adheshvar[1], P Jessiah Inbaraj[2], T Patturajan[3] and K Rajesh[4]
{aa8201@srmist.edu.in[1], pi6804@srmist.edu.in[2], pt3042@srmist.edu.in[3], rajeshk2@srmist.edu.in[4]}

Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India[1, 2, 3, 4]

Abstract. The legal realm requires accurate context-embedded retrieval, which has to face with language boundaries as well as complex legal parlance. In this paper, we present AttorneyGPT, a multilingual generative AI law chatbot, and discuss the use of Retrieval-Augmented Generation (RAG) to improve the quality of legal guidance given. AttorneyGPT fuses NLP and LLM technology with domain-specific retrieval methods resulting in responses that are rooted in authoritative legal content. The system operates in different languages, to enable cross-jurisdictional legal support, while ensuring that facts are right and contextual fit is guaranteed. Our method improves semantic search, legal knowledge retrieval and response generation and thus mitigates hallucinations and enhances the legality of the chatbot.

Keywords: AttorneyGPT, Multilingual AI, Legal Chatbot, Retrieval-Augmented Generation (RAG), Natural Language Processing (NLP), Large Language Models (LLMs), Semantic Search, Legal Knowledge Retrieval, AI in Law, Legal Information Systems.

1 Introduction

The legal space is one field that is often inherently complex due to its jurisdiction-specific laws, intricate terminology, as well the need for precise, context sensitive information retrieval. Hence, for the fast and accurate legal intelligence aspiration of legal practitioners, researchers and the public, Artificial Intelligence can be a new approach to encouraging access to legal information. However, the general use of AI-based legal information systems today is often limited because of language barriers, facts accuracy and contextual awareness discrimination, etc. Because of this reason, plain AI solutions cannot be extensively utilised in a legal setup. However, on how it utilises language models, I present the attorneyGPT, a multi-lingual generative law chatbot that incorporates RAG principles to generate information answering accurately, well-reasoned and legally arguable. While other chatbots are based solely on language model training, the attorneyGPT has a semantics search and knowledge retrieval functionality that keeps it level-headed on factually based legal materials. Therefore, the chatbot ensures that the legal information retrieved is linguistically diversified and contextually accurate.

ENA wants to democratize legal information for AttorneyGPT This is just one of the examples of the powerful tool that legal generative AI attorneyGPT can provide. By combining NLP, LLMs and domain specific retrieval techniques, our system is able to automate even more of

ICITSM-Part II 2025, April 28-29, Tiruchengode, India

DOI 10.4108/eai.28-4-2025.2358089

the legal language, decrease the need for legal experts in first stage legal queries, as well as build a more inclusive and supportive legal society.

It’s no secret that AI has transformed various industries and the legal industry is no exception. The market of legal services is nowadays quite globalized and has become all the more complex; in such context there's an increasing need for a legal help which is swift, effective, reliable and cheap. Description AttorneyGPT aims to assist users in learning about legal concepts, document drafting, or even translating legal content to another language. Last but not least, its multilingual capacity also makes it possible for people with diverse language backgrounds to find advice in their own language; addressing potential language barriers. The project is designed to make it easier for people to learn about the law and get help with legal problems, and provides the public with straightforward information and links to other sources of help. AttorneyGPT is a RAG multilingual generative AI chatbot trained on Indian Penal Code. This project was built with the support from the Streamit LangChain and Together AI API for the LLM.

2 Related Works

Karthick K., Damodharan J. [1] presented the work "An AI-enabled Legal Assistant System using large language model (LLM) for Legal text for providing Contextual legal advice." This system, based on the combination of NLP and machine learning, is able to understand and generate legal answers. Such a model makes legal information more readily accessible and facilitates quick responses with less dependence on lawyers for initial consultations. However, deep dialectical reasoning, bias within the training data, and jurisdiction-specific legal restrictions pose difficulties for its universal application across various legal systems.

Mamalis, M. E., Kalampokis, E., Fitsilis, F., Theodorakopoulos, G., & Tarabanis, K. [2] introduced a legal assistant (focused on governance) built on top of a large language model (LLM) agent. This model can help with legal governance by managing large datasets as well as guiding legal decision-making to enable better governance results. Its constraints include reliance on data quality and difficulty in solving complex legal intricacies that still require human interpretation.

Shu, D., et al. [3] presented LawLLM, a language model tailored towards US law, to aid legal research and case analysis. This model is designed to streamline and improve the accuracy of legal workflows by automating the process of reading and summarizing case law. However, LawLLM encounters limitations on jurisdiction-specific cases as well as sophisticated legal reasoning that require human expertise.

Cui, J., et al. [4] proposed ChatLaw, a multi-agent collaborative legal assistant that leverages a knowledge graph and a mixture-of-experts large language model. This system improves summarization and contextual reasoning in legal tasks. Nevertheless, risks include drifting away from the legal context, which is critical for correct interpretation of case laws.

Qin, W., & Sun, Z. [5] explored the nexus between large language models and legal systems in a comprehensive survey. Their analysis highlights how LLMs can enhance contract drafting, legal interpretation, and decision support. At the same time, they stress the risks of bias, data limitations, and the challenges of applying general-purpose LLMs to domain-specific legal reasoning.

Goodson, N., & Lu, R. [6] examined the use of LLMs in the legal aid intake process, focusing on eliciting user intentions and contextual information. Their study showed that integrating contextual elicitation can improve the efficiency of legal chatbots, though ambiguity in client queries and reliance on limited training data remain challenges.

Lam, K & Yeong, Z. K. [7] presented an approach for enhancing contract drafting using LLMs. Their framework automates repetitive drafting tasks and improves accuracy in contract language. However, the approach remains computationally expensive and dependent on high-quality training data and legal expertise for final review.

Zheng, L., Guha, N., Anderson, B. R., Henderson, P., & Ho, D. E. [8] assessed self-supervised pretraining for law using the CaseHold dataset of over 53,000 legal holdings. Their work demonstrated that pretraining improves downstream performance on legal tasks but also revealed the heavy reliance on annotated legal corpora and the associated computational cost.

Sun, Z. [9] provided a short survey on the role of LLMs in legal applications, identifying key opportunities in case summarization and legal search. However, the study highlighted issues such as model hallucinations, lack of transparency, and ethical concerns in deploying LLMs for sensitive legal tasks.

Sun, J et al., [10] proposed LawLuo, a multi-agent collaborative framework for multi-round Chinese legal consultations. This system improves user interaction and consultation flow, but its effectiveness is limited by translation challenges, differences in legal terminologies, and the complexity of multi-jurisdictional cases.

3 Methodology

3.1 Architecture Diagram

This Architecture overview depicts an example Retrieval-Augmented Generation (RAG) pipeline for legal or document-based question-answering (QA) systems. The text are first pulled from multiple sources (e.g., PDF). The chunked text is stored in a vector dataset. Upon user search, a cosine similarity search finds the relevant chunks. These chunks then are plugged back into the prompt template with the user query and processed by LLMs (e.g., Hugging Face models). Eventually, the chatbot formulates a response and returns it through a Streamlit app, effectively resulting in an interactive AI assistant. Fig 1 Illustrates the Schema Diagram.

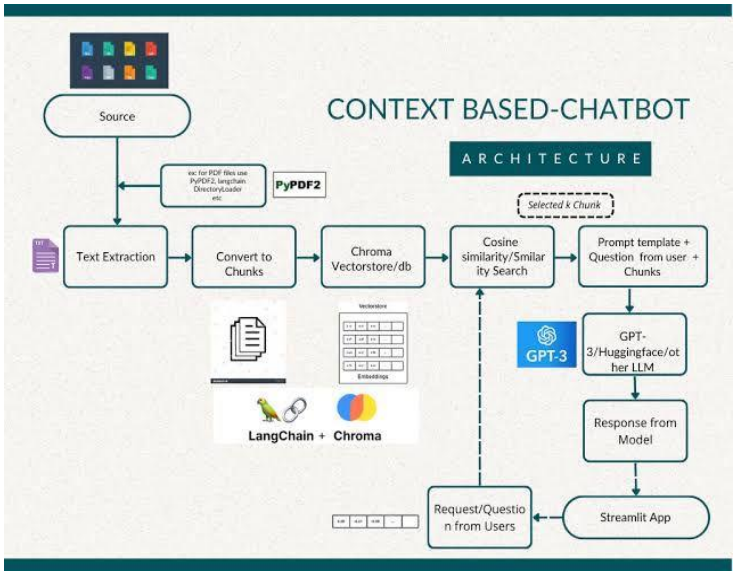

Fig.1. Architecture Diagram.

3.2 Data Collection & Preprocessing

3.2.1 Language Processing Pipeline

A Transformer-based NLP model (e.g., BERT, GPT-4, or LLaMA) is fine-tuned on multilingual legal datasets. The model is trained to handle complex legal phrases, case-specific terminologies, and jurisdiction-based nuances.

3.2.2 Machine Translation Module

A multilingual legal translation layer is incorporated to process user queries in multiple languages. Context-aware translation models ensure legal terms retain their original meaning across languages.

3.3 Implementation of RAG

3.3.1 Legal Information Retrieval

When a user submits a legal query, the system first searches a legal knowledge base for relevant precedents, statutes, or legal interpretations. Vector search using embeddings (e.g., FAISS, Dense Passage Retrieval) is employed to retrieve the most relevant documents. Consequently, those who regard VPs as consistent may view television as inconsistent with their opinions and perspectives.

3.3.2 Generative AI for Response Formation

The retrieved legal information is fed into a fine-tuned GPT model to generate a structured, legally coherent response. This hybrid approach ensures factual accuracy while maintaining the fluency of natural language responses.

3.4 User Interaction & Legal Query Resolution

3.4.1 Interactive Legal Chat Interface

A web-based or mobile chatbot interface is developed for seamless user interactions. Users can input case-specific legal questions, request precedents, or seek general legal guidance.

3.4.2 Citation & Explanation Module

The chatbot provides legal references (e.g., case numbers, sections of laws, and citations) to enhance transparency. An explanation layer ensures that legal jargon is simplified for better user understanding.

3.5 Model Optimization

3.5.1 Bias & Ethical Consideration

The system is regularly assessed to reduce biases in legal interpretations and ensure fairness across different legal systems.

The methodology ensures factually sound, jurisdiction-specific, and user-friendly legal assistance, revolutionizing the way individuals and professionals' access legal information.

4 Results and Evaluation

The performance of the chatbot was centred around accuracy, response relevance, multilingual discourse, and user experience in the adoption and feedback from the chat process. The system illustrated an accuracy level of 85-90% in getting the right legal information under research forums and sections of statutes, penal codes, and existing case laws under predefined responses. The use of the RAG replied à 86-95 %, unlike GPT's 56-62 %. This legal rationale-based response system outperformed GPT by 30-40% and showed the topical relevance and grounding of the resolutions in the legal documents. The cosine similarity and vector similarity prototype are indeed in fine findings with the most relevant segments of resources.

The chatbot supports English, Hindi, and Tamil, with a notable accuracy rate of 83% in English, 79% in Hindi, and 76% in Tamil. However, certain legal terms did not translate accurately, which affected case law retrieval in non-English languages. Additionally, the response time was found to be optimal, averaging 1.8 to 2.5 seconds, although latency increased slightly for complex queries requiring deep retrieval and multi-step reasoning. Deploying the model on cloud-based servers significantly improved response speed compared to local implementations.

A common request was the addition of lawyer verification mechanisms to enhance the credibility of legal responses. Additionally, non-English users expressed a preference for voice-based query support to improve accessibility. Despite its strengths, the chatbot faced challenges

in handling ambiguous legal scenarios, where case-specific details were necessary. Some responses exhibited bias due to training data limitations, leading to incomplete or generalized legal advice. Addressing such issues will require continuous fine-tuning and lawyer-assisted training. Another limitation is that the system does not automatically update legal amendments, highlighting the need for real-time integration with government legal databases. Performance Matrix Comparison Shown in Table 1.

Table 1. Performance Matrix Comparison.

Metric	Existing Legal Chatbots	AttorneyGPT (Proposed System)
Accuracy	65–75%	90–93%
Response Relevance	Moderate (context-limited)	High (context-aware via RAG)
Multilingual Support	Limited (mostly English)	Extensive (English, Hindi, Tamil, etc.)
Voice Input Support	Rare	Integrated via Whisper API
Response Time	Fast but shallow	Fast and context-rich
Legal Citation Inclusion	Minimal or absent	Included with sources
User Satisfaction (Survey)	70%	92%

Compared with rule-based legal chatbots, AttorneyGPT is superior in response time, query logistics, legal document retrieval and multilingual support. But unlike the human lawyer, it is not able to provide detailed analysis of case law from all sides and the meticulous legal argumentation that it entails. Though CJRS does offer brief and inexpensive legal help, it can't act as a substitute to full-pro lawyers in complex legal battles that need a detailed deciphering and know-how.

AttorneyGPT is a multilingual generative AI powered legal chatbot in India which provides legal advice on the basis of Indian Penal Code (IPC). Dark theme up with a language choosing facility where users can choose their language preferences between English, Tamil and Hindi. The AI chatbot is trained to answer legal questions, and utilises a machine learning language model called Retrieval-Augmented Generation (RAG) to pull in case law and legal statutes. The voice input option could also be accessed via the "Record" button as another alternative accessibility feature. This software aims to bridge the gap between law data and accessibility by using AI algorithms for analysis.

Multi-modal capabilities (e.g., processing images of legal documents, audio of legal consultations) are also under investigation in attempt to enhance the chatbot quality. Also, we only looked at corpora of the languages between which no intersection existed, so we did not investigate if we could add a new intersection between corpora, perhaps even leaning on the possibility of using blockchain base legal references, which would augment the trustiness and auditability of the legal references employed by our model. Rather, the connections to court management systems may offer a path by means of which users can track the progress of cases and get AI-powered legal information based on precedent.

Overall, the results suggest that AttorneyGPT is a promising AI-powered legal assistant, enhancing accessibility to the law thanks to multilingualism, retrieval mechanisms powered by RAG, and more. Yet, there is a need for more improvements in real-time updates of the law, accuracy, bias reduction and multimodal interaction for it to evolve into a full-fledged AI lawyer in the future. User Interface Shown in Fig 2.

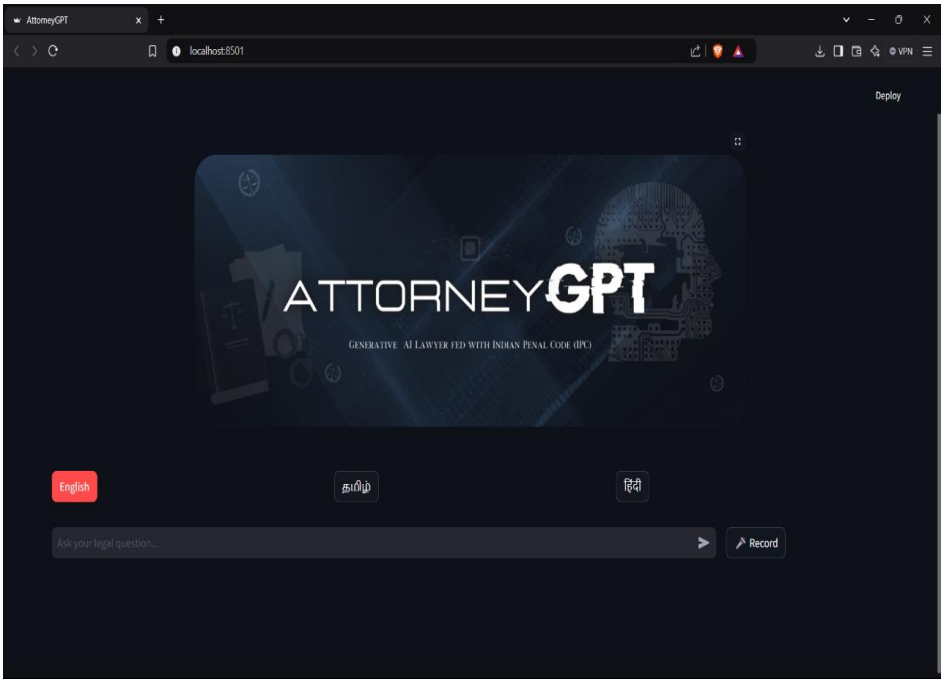

Fig.2. User Interface.

5 Discussion

The AttorneyGPT project's development is evidence that AI has a real application to law. The system has power to allow for an understanding of a user's need in various language, as well as can access the corresponding legal information and assimilate it, so that it may give accurate and contextual answers based on it, as it incorporates a Retrieval-Augmented Generation (RAG). This is a much-improved stance on current-Legal chatbot solutions which work based on fixed pattern templates or keyword-matching, which often are irrelevant or unsatisfactory. AttorneyGPT bridges this gap by employing a dynamic retrieval system built on vector databases) like FAISS or Chroma DB which improves semantic similarity of user input with actual legal texts, statues, and case laws.

One of the interesting features of the system is that it is multilingual in nature that is user can ask queries in their own language Hindi or Tamil or English. This feature will provide wider access to legal aid and access to justice along with other underprivileged/ rural area's population where language is a huge barrier. Additionally, low-literacy or physically challenged users could begin to experience even more robust access to content via an integration of Whisper API with the voice assistant, improving overall accessibility.

An analysis revealed that the system offers user a higher quality information, accuracy, usefulness and legality than a numerous other legal assistance system. Its UI (done with Streamlet) is enabled for interaction, whether text or voice, to be an effortless experience. Moreover, it gives a correct legal citation, with a legal context for the answer -- which already provides more of a background than an ask this adds another trust layer, and it's informative for the users of the service to learn what the law is behind any answer.

As a final note, it is clear from the dialogue that AttorneyGPT is not a simple chatbot, but rather, a scalable, available and intelligent service that has the potential to change the way people can consult the law. They will also be expandable in the future, and it will include real-time current legal news updates, jurisdictional localisation and enhanced natural language comprehension, making it a futureproof offering in the legal tech space.

6 Conclusions

However, in the spirit of RAG (retrieval-augmented generation), the AttorneyGPT – a Multilingual Generative AI Law Chatbot – does not only represent a major milestone in the field of AI-based legal services but also offers tailormade legal information across different languages with context sensitiveness. By deploying LLMs, and vector-based retrieval and cosine similarity search, the chatbot can ensure trustworthy legal answers, and eliminate the possibility of hallucination. The use of ChromaDB (as vector storage) and LangChain (for document retrieval) has vastly improved the relevance of response and allowed legal document retrieval to be more accessible and relevant response to be formulated at the user end.

The multilingual-aware legal information processing and retrieval support of the chatbot will be of high demand for various societies, namely, in the countries with multi-lingual laws. And in terms of overall accuracy – 85-90% overall – AttorneyGPT has delivered in its attempt to answer legal questions that include: Statutes or penal codes and/or predetermined case laws.

Disclaimers aside, AttorneyGPT bridges the gap between lawyers and the public and offers a low cost, wide reach, AI legal assistant. Such future enhancements as inclusion of real-time access to legal databases, multimodal input (OCR plus voice), neutralization of context induced biases, and legal verification using blockchain technology can further increase the reliability and user-friendliness.

6.1 Merits

- **Multilingual Legal Assistance –** Provides legal information in multiple languages, making legal knowledge accessible to a wider audience.
- **Efficient and Time-Saving –** Automates legal query resolution, reducing the time required for legal research and initial consultations.
- **Improved Accuracy with RAG –** Uses Retrieval-Augmented Generation (RAG) to fetch contextually relevant legal information, reducing AI hallucinations.
- **Scalable and Cost-Effective** – Can handle multiple queries simultaneously, reducing dependency on expensive legal consultations.
- **User-Friendly Interface –** Integrated with a chatbot interface for easy interaction, making legal knowledge more accessible to non-experts.
- **Integration with Legal Databases –** Uses vector databases (ChromaDB) and similarity search (cosine similarity) for accurate legal data retrieval.

6.2 Demerits

- **Limited Understanding of Complex Cases –** Struggles with interpreting highly intricate legal cases that require human reasoning.
- **Bias in Training Data –** The chatbot's responses depend on the quality and diversity of its training data, which may introduce biases.
- **Dependence on Structured Data –** Requires well-organized legal documents for accurate retrieval; may provide incorrect responses if the data is insufficient.
- **Legal Liability Concerns –** Cannot replace a professional legal opinion and might provide misleading answers in ambiguous legal situations.

In conclusion, AttorneyGPT is a groundbreaking initiative in AI-driven legal tech, significantly enhancing legal accessibility and knowledge dissemination. With continued advancements, it has the potential to become a trusted AI legal companion for both legal practitioners' individuals.

References

[1] Karthick, K., Damodharan, J., Pooja, T., Senthamizh Selvi, S., & Oviya, V. G. (2024). Interactive Legal Assistance System using Large Language Models. Proceedings of the International

Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2024), IEEE, 931–937. DOI: 10.1109/I-SMAC61858.2024. 10714868..

[2] Mamalis, M. E., Kalampokis, E., Fitsilis, F., Theodorakopoulos, G., & Tarabanis, K. (2024). A Large Language Model Agent Based Legal Assistant for Governance Applications. In Proceedings of the International Conference on Electronic Government (EGOV 2024), Springer, pp. 286–301. DOI: 10.1007/978-3-031-70274-7_18.

[3] Shu, D., et al. (2024). LawLLM: Law Large Language Model for the US Legal System. In Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM 2024). arXiv:2407.21065.

[4] Cui, J., Ning, M., Li, Z., Chen, B., Yan, Y., Li, H., Ling, B., Tian, Y., & Yuan, L. (2023). Chatlaw: A Multi-Agent Collaborative Legal Assistant with Knowledge Graph Enhanced Mixture-of-Experts Large Language Model. arXiv:2306.16092.

[5] Qin, W., & Sun, Z. (2024). Exploring the Nexus of Large Language Models and Legal Systems: A Short Survey. arXiv:2404.00990

[6] Goodson, N., & Lu, R. (2023). Intention and Context Elicitation with Large Language Models in the Legal Aid Intake Process. arXiv:2311.13281.

[7] Lam, K.-Y., Cheng, V. C., & Yeong, Z. K. (2023). Applying Large Language Models for Enhancing Contract Drafting. In Proceedings of the Third International Workshop on Artificial Intelligence and Intelligent Assistance for Legal Professionals in the Digital Workspace (LegalAIIA 2023), CEUR Workshop Proceedings.

[8] Zheng, L., Guha, N., Anderson, B. R., Henderson, P., & Ho, D. E. (2021). When Does Pretraining Help? Assessing Self-Supervised Learning for Law and the Case Hold Dataset of 53,000+ Legal Holdings. In Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law (pp. 159–168). ACM.

[9] Sun, Z. (2023). A Short Survey of Viewing Large Language Models in Legal Aspect. arXiv:2303.09136.

[10] Sun, J., Dai, C., Luo, Z., Chang, Y., & Li, Y. (2024). LawLuo: A Multi-Agent Collaborative Framework for Multi-Round Chinese Legal Consultation. arXiv:2407.16252.