

AI-Based Legal Chatbot for Providing Basic Legal Assistance

Tanmay Singh Panwar

Software Engineering

Vivekananda Global University

Raghav Sharma

Law

Vivekananda Global University

Prashant Jaiswal

Software Engineering

Vivekananda Global University

Mansha Kumari

Management

Vivekananda Global University

Prerit Kumbhar

Forensic Science

Vivekananda Global University

BTech (Computer Science & Engineering)

1. ABSTRACT

Honestly, getting legal help is a headache for most people - I've noticed this in my family as well. Some time ago, my uncle got a legal notice but didn't understand what it meant or what to do next. This situation is very common in India. Many people have minor or simple legal problems but do not know where to turn for help. Hiring an attorney can be expensive, and free legal services are often too far away or too crowded.

This project is based on this problem. As part of my Bachelor of Technology (BTech) studies, I built a chatbot that tries to make legal information accessible. Chatbots use plain language to understand what people are asking and then provide clear and understandable answers.

I use Python to build the backend and libraries like spaCy and NLTK to process the script. For questions that the system does not recognize, connect to the LLM API to check the answer.

During the inspection, the chatbot can answer a variety of common legal questions, such as tenant issues, payroll issues, and consumer complaints. Of course, it is not replacing the lawyer, and this is obvious to the client. Rather, it helps people understand their situation and decide whether they need professional legal help.

Even this small step can make a big difference.

2. INTRODUCTION

Imagine a situation where someone you know has received a summons or argued with a seller or landlord. Many people get scared. Not that it's a huge problem, but legal language is very complex and the whole process is confusing and intimidating. This is a real problem in our country.

India has a large population of over 1.4 billion people, but the number of lawyers compared to the population is very low. Even if lawyers are available, most people don't bother to ask for help. There is free legal assistance, such as that offered by the National Legal Services Authority, but many people are unaware of this possibility or have difficulty accessing it, especially in smaller towns and villages.

Reading other research articles on this topic, I understood that technology can help in this regard. For example, studies conducted by Feng (2018) and Robert Dale (2019) showed how natural language processing and chatbots can be used in the legal sector.

This made me realize that chatbots can help, not replace lawyers, but help people. He can explain their problem and give them an idea of what they can do.

This project is an attempt to build something like that. The chatbot speaks in simple language, uses a familiar font, and can be used by anyone with a smartphone—no legal training required.

3. LITERATURE REVIEW

3.1 Existing

When I started researching this topic, I was surprised by how much work had already been done. Most existing systems are for lawyers and forensic investigators.

One of the most discussed examples from the past is a chatbot called DoNotPay. It was later expanded to include unsubscription and handling false accusations. [2] Although it was initially based on law, legal chatbots do actually work in practice.

The research area includes ROSS Intelligence, a system built using IBM Watson. But it is useful. Only useful for professionals. The gap for non-professional users is still large.

3.2 Rule-based and NLP-based chatbots.

When reading the work of Quaresma and Goncalves (2006), I found a very useful classification of types of chatbots. If the user says X, they are predictable and accurate in answering specific questions. Even a small change in terminology can disrupt the entire process.

NLP-based chatbots are more flexible. They try to understand the meaning of words, such as intent detection and object extraction. They are not perfect, but they cope well with a wide range of languages. Most modern systems use a combination of both methods.

3.3 Large language models change the rules of the game.

The most useful part of my overall research is on transducer-based models such as BERT, GPT, and similar systems. These models are trained on large amounts of documents, and their results are predicted. Context and follow-up questions can generate explanations that actually make sense to non-experts. This is important for legal chatbots. This means that the system can explain what “very aggressive” means in English, rather than providing a definition. As Brown et al. (2020) show, the problem with LLMs is that they sometimes seem promising but give wrong answers. This is a major concern in legal contexts. Therefore, a combination of methods (predetermined answers and reverse LLM) makes more sense to me than relying on LLM alone.

4. PROBLEM STATEMENT

The main problem can be stated simply: most ordinary people do not know their basic rights and do not have easy access to this information. For example, when a landlord refuses to return a deposit, an employer withholds a monthly salary, or a retailer sells a defective product, people either choose to ignore it or spend a lot of time and money trying to figure out what to do.

Practical solutions usually require users to have extensive legal knowledge, are very expensive, or are simply not available at 2:00 a.m. - precisely when panic sets in after receiving a legal summons. Therefore, there is a clear need for a tool that is free, available 24/7, uses plain language, and allows users to clarify the nature of the situation they are facing.

The chatbot proposed in this project represents an attempt to solve this problem. While it cannot guarantee victory in court, it can provide users with information about the specific legal situation they are facing, indicate which rules may apply, and determine the main course of action. Indeed, this initial step - understanding the essence of the situation - often turns out to be the most difficult.

5. PROPOSED SYSTEM

5.1 System Architecture

The system consists of three main components. First, there is the user interface: a simple browser-based chat window where the user enters their query. Second, there is a processing layer that receives the input and tries to interpret the user's intent. Third, there is a response layer that either selects an answer from a compiled FAQ database or, if there is no ready-made answer, sends the query to the LLM API.

Currently, the entire system is hosted locally, with the Flask framework handling the operations on the server. The front-end is made up of simple HTML and JavaScript code: nothing too complicated here; it is purely functional.

5.2 Data Flow: Step by Step

When a user enters a query, such as “My boss hasn't paid me for two months, what can I do?”, the system first cleans the text (removes stop words and performs tokenization). Then it checks whether the query matches a specific intent stored in the database. If a match is found, the specified response is returned. If there is not enough correspondence, the request will be sent to the LLM API along with a system instruction stating: “Please provide the relevant legal

information in clear and understandable language, and in case of serious problems, it is always recommended to consult a lawyer.”

5.3 Technologies used

- Python 3.10 - main programming language
- spaCy and NLTK - for natural language processing (NLP) tasks such as tokenization and intent matching
- Flask - server
- LLM API (Antropic / OpenAI) - for handling complex or open-ended queries
- Conversational interface (frontend) HTML, CSS, JavaScript
- JSON file - for storing the FAQ database and the intent representation of the responses

6. METHODOLOGY

6.1 Data Collection

Creating the Frequently Asked Questions (FAQ) database took longer than I had anticipated. I analyzed the NALSA website, the consumer court guidelines hosted on the Ministry of Consumer Affairs' website, and selected pages from the Indian Kanoon platform. I also examined FAQs found across various judicial forums. In total, I gathered 150 question-and-answer pairs covering the fundamental aspects of property disputes, employment law, consumer issues, and family law.

The data was manually curated to eliminate legal jargon wherever possible, to rewrite the answers in simple language, and to assign a specific topic category and several keyword variations to each entry.

6.2 Methodology

As I mentioned earlier, this system employs a hybrid approach. For questions phrased in simple, straightforward terms, it utilizes a keyword and intent-matching mechanism. For questions involving more complex or unconventional phrasing, it leverages an LLM. This design ensures that the system remains robust for standard FAQs while simultaneously proving effective in more intricate scenarios.

6.3 Tools and Configuration

The entirety of this project was developed on a local Windows machine using VS Code. Prior to implementing the plugin, I used Postman to test the API endpoints. The spaCy model utilized for NLP is the English Small Model ('en_core_web_sm'); it is lightweight and well-suited for basic intent classification.

7. FEATURES OF THE SYSTEM

- Legal Answers: Handles questions related to employment, property, consumer rights, and family law.
- Plain Language Answers: No complex legal language - everything is explained in the same way you would explain it to a friend.
- Built-in Disclaimer: Each answer includes a note reminding the user that the information provided is for informational purposes only and does not constitute professional legal advice.

- **Context Retention:** Within a session, the chatbot remembers previous questions, ensuring that subsequent questions remain relevant and meaningful.

8. ADVANTAGES AND LIMITATIONS

8.1 Effectiveness

Its biggest advantage is its accessibility. Anyone - anywhere, anywhere - can use this chatbot completely free of charge. Users don't need to know any rules or guidelines before they start using it. For someone living in a small town who has received a notice from their employer about their termination and is unsure whether it is legal, even the basic instructions provided by this chatbot can be of great importance.

Thanks to its hybrid architecture, the chatbot doesn't "get stuck" when users ask questions that go beyond the limits of its predefined answers (FAQ). Instead of displaying a useful "I don't understand" message, it smoothly switches to its main Large Language Model (LLM).

8.2 Observed Limitations

The most obvious of these is that Large Language Models (LLMs) can sometimes give wrong answers, and they are quite confident in their accuracy. While this is acceptable for general questions, it is a serious concern when legal issues arise. To minimize risk, I tried to implement very precise procedures to limit the scope of LLM; however, this is not a solution that guarantees 100% security.

Furthermore, this program may not handle such complex or complicated legal situations, where the appropriate response depends on many factors - such as the jurisdiction of the court, the terms of the specific contract, or criminal history. Furthermore, given that it currently only works in English, its usefulness to a large segment of the Indian population is very limited.

9. FUTURE SCOPE

This project opens several potential avenues for development. The most obvious option is to directly integrate the chat server with the legal database - like the Indian Kanoon API. This allows the chat server to take advantage of the latest case law when formulating its responses, making the responses more reliable and practical.

I also plan to introduce support for Hindi. In India, millions of people find it much more comfortable to communicate in Hindi than in English, but legal awareness within these communities is still relatively low. Even a simplified version of this chat server could reach countless individuals who currently lack access to legal information.

In the long term, an attractive improvement is to implement a document analysis feature, which allows users to upload legal documents or contracts for the chat server to analyze and explain in simple, accessible language. From a technical perspective - thanks to current advances in legal AI - such a solution is already feasible and proving very useful in everyday situations. If public legal aid agencies publicly adopt such a system - for example, by enabling a

chatbot to connect users with nearby legal aid centers or helping them file complaints online – it grows into a tool with even greater impact and transformative potential.

10. CONCLUSION

The project grew out of a simple observation: legal aid is still inaccessible to many ordinary people—a problem that technology can address to some extent. The chatbot developed as part of this project is not a comprehensive solution; It has real-world limitations and certainly does not act as a substitute for a qualified legal expert. However, it can achieve one significant thing: it can give people a starting point.

Understanding the nature of your own legal issues, knowing the laws that apply, and being informed about your rights in certain situations are of the utmost importance. These factors can determine whether a person decides to fight for their rights or succumbs to an unfair situation.

From a technical perspective, the project demonstrates how Natural Language Processing (NLP) and Large Language Model (LLM) tools can be integrated in practical and easily accessible applications. The hybrid approach used proved to be effective during testing, and the system successfully handled many common queries. With further improvements in areas such as multilingual support, real-time data integration, and document clarification, the system can turn into an invaluable tool for raising legal awareness in India.

11. REFERENCES

- [1] A. J. Sandefur, "Access to What?" *Daedalus*, vol. 148, no. 1, pp. 49–55, Jan. 2019.
- [2] J. Brownsword, E. Scotford, and K. Yeung, *The Oxford Handbook of Law, Regulation and Technology*. Oxford, UK: Oxford University Press, 2017.
- [3] N. Aletras, D. Tsarapatsanis, D. Preoiuc-Pietro, and V. Lampos, "Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective," *PeerJ Comput. Sci.*, vol. 2, p. e93, 2016.
- [4] Y. Feng, "Legal question answering using structured knowledge and neural networks," in *Proc. ACL 2018 Student Research Workshop*, Melbourne, Australia, Jul. 2018, pp. 26–32.
- [5] T. Brown et al., "Language models are few-shot learners," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, 2020, pp. 1877–1901.
- [6] M. Quaresma and T. Goncalves, "Using NLP techniques in a question/answer system in the legal domain," in *Proc. 1st Workshop on Automated Legal Text Processing (ALTP-2006)*, 2006, pp. 1–8.
- [7] R. Dale, "Law and word order: NLP in legal tech," *Nat. Lang. Eng.*, vol. 25, no. 1, pp. 211–217, 2019.