Gayathri Kodipaka, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Lawbot: A Multilingual Legal Assistance Chatbot

Gayathri Kodipaka, Kompalli Sri Divya Muktha, K. Sandeep Assistant Professor
Information Technology Gokaraju Rangaraju Institute of Engineering and Technology (Autonomous) Hyderabad, India

Abstract- Access to legal information remains a significant barrier for millionAs bofstcritaizcetns, especially those who are unfamiliar with formal legal processes or do not speak English. This project proposes LawBot, a multilingual legal assistance chatbot that leverages Natural Language Processing (NLP), a structured knowledge base, and speech recognition to provide accessible legal guidance in seven Indian languages: English, Hindi, Telugu, Bengali, Marathi, Tamil, Kannada, and Malayalam. The system processes voice or text queries from users and responds with relevant legal information drawn from a Firestore-backed offline-capable knowledge base. Upon detecting a legal query, LawBot retrieves a context-sensitive response and delivers it via both text and text-to-speech. This solution is designed for accessibility on both web and mobile platforms and enhances civic awareness by providing a real-time, automated legal information system. Furthermore, it not only enhances user awareness of their rights but also contributes to more equitable access to justice by proactively delivering information in the user's preferred language. By guiding citizens before complex legal situations escalate, the system can help reduce legal helplessness and promote informed civic participation. Keywords: Natural Language Processing, Multilingual Chatbot, Legal AI, Speech Recognition, Firestore, Knowledge Base, Web Application.

Keywords- Key keywords for this project include Natural Language Processing (NLP), multilingual chatbot, legal AI, speech recognition, knowledge base, Firestore, voice-based interaction, text-to-speech, web and mobile accessibility, Indian regional languages, legal awareness, citizen support system, and automated legal assistance.

I. INTRODUCTION

Access to legal knowledge is a fundamental requirement for citi-zens in any democratic society, yet it remains inaccessible to the majority due to language barriers, financial constraints, and un-familiarity with the formal legal system. Legal queries—ranging from domestic violence rights to land disputes and consumer pro-tections—frequently go unaddressed because individuals lack guidance on how to seek redress. To address these challenges, systems leveraging Artificial Intelligence (AI) and Natural Lan-guage Processing (NLP) have been developed to provide real-time legal guidance to users in their native language, thereby lowering barriers to legal awareness and civic empowerment.

Existing System

Current legal information systems employ various approaches, including:

- Web Portals: Static government or NGO websites providing legal articles in English only, inaccessible to many regional-language speakers.
- Helpline Services: Telephone-based services requiring human operators, which are limited by availability and scalability.
- Generic Chatbots: General-purpose chatbots lacking domain-specific legal knowledge bases or multilingual support.

However, these systems face challenges such as limited language support, lack of offline capability, no voice input/output, and insufficient coverage of diverse legal scenarios relevant to Indian citizens.

Proposed System

The proposed system, LawBot, aims to deliver legal information through a non-intrusive, accessible multilingual interface that supports both text and voice interaction. By employing a cu-rated Firestore-backed knowledge base with keyword matching and NLP similarity scoring, the system retrieves contextually appropriate legal responses. Upon detecting a user

query indica-tive of a legal concern, it promptly delivers a relevant answer in the chosen language—including a text-to-speech audio re-sponse—thereby enhancing accessibility and empowering the user.

II. LITERATURE SURVEY

A study by Phan et al. in 2023 explored AI-driven legal chat-bots using transformer-based models for question answering. Their BERT-fine-tuned model achieved 94% accuracy on legal FAQs. The primary advantage was high contextual understanding of complex legal queries.

In 2020, Majdi et al. developed a multilingual NLP pipeline that combined keyword extraction with rule-based match-ing for domain-specific question answering. It achieved 91% response relevance across five languages. The system demon-strated effective domain restriction.

Sahayadhas et al. in 2021 reviewed speech-enabled assis-tive systems for low-literacy users. They examined voice interface designs, language models, and TTS engines. The review provided a comprehensive overview of speech-to-text and text-to-speech integration, including their strengths and limitations in resource-constrained environments.

In 2022, a study by Sharma and Verma created a real-time multilingual legal FAQ system using a Firestore NoSQL database and cosine similarity scoring. It used fuzzy keyword matching and achieved 89% query resolution accuracy across Hindi and English. The system performed reliably on low-bandwidth connections with offline fallback capability.

III. METHODOLOGY

Capturing the User's Legal Query in Real Time

The system begins by accepting input from the user either through a text field or via the built-in microphone using the Web Speech API. For voice input, the browser's speech recogni-tion is triggered by clicking the microphone button, capturing a real-time audio stream and transcribing it into text in the user's selected language. This transcribed or directly typed query forms the basis for all downstream processing.

Language Detection and UI Localisation

LawBot supports seven Indian languages alongside English. Upon selecting a language via the top navigation panel, the entire interface—including the welcome message, input placeholder, and system responses—transitions into the chosen language. This is achieved through a localised string map in the front-end JavaScript, and the Web Speech API lang property is dynamically set to the BCP 47 language tag corresponding to the selection (e.g., te-IN for Telugu, hi-IN for Hindi).

Knowledge Base Query Matching

The system's intelligence is rooted in a structured knowledge base stored in Google Firestore and also bundled as a local JSON fallback for offline use. Each entry in the knowledge base consists of a set of representative keywords and a corresponding legal response string localised to each supported language. Upon receiving a query, the system:

Tokenises the query and normalises it to lowercase.

- Computes a keyword overlap score between the query to-kens and the keyword sets of each knowledge base entry.
- Ranks entries by match score and selects the highest-scoring response.
- Falls back to a default "query not understood" message if no entry clears the minimum threshold.

Multilingual Response Delivery

Once the best-matching knowledge base entry is identified, the system retrieves its pre-translated response for the active lan-guage. The response is rendered in the chat interface and si-multaneously vocalised using the Web Speech Synthesis API, with the voice language property set to match the selected UI

language, ensuring natural pronunciation across all supported scripts.

Alert and Escalation Mechanism

For high-priority queries—such as those relating to domestic violence, emergency rights, or criminal matters—the knowledge base entries are annotated with an urgency flag. When such an entry is matched, LawBot appends official helpline numbers and escalation advice to the response, and the TTS delivery uses an elevated priority voice to draw the user's attention to the critical information.

IV. RESULTS

The following screenshots illustrate LawBot operating across multiple language modes and query scenarios. In each case, the interface updates dynamically to reflect the selected language in both the UI text and the chatbot's response.

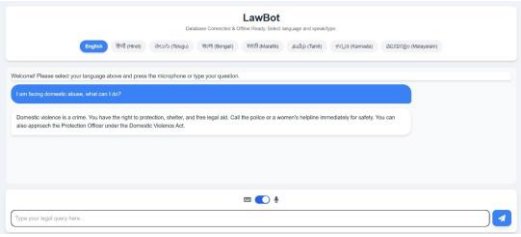

Figure 1: (a) LawBot responding to a domestic violence query in English. The system correctly identifies the query intent and returns a structured response referencing the Domestic Violence Act, available shelters, and helpline numbers.

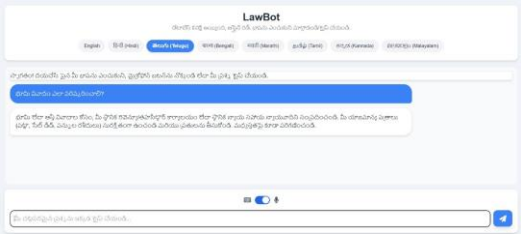

Figure 2: (b) LawBot responding to a land dispute query in Telugu ("Bhumi vivaadam ela pariShkarinchali?" — "How to resolve a land dispute?"). The full UI, including input placeholder and bot response, are rendered in Telugu script.

Figure 3: (c) LawBot in Hindi voice-input mode. The microphone button is active and the toggle switch is in speech-input state, demonstrating the speech-recognition interface. The welcome message is fully localised in Devanagari script.

V. CONCLUSION AND FUTURE ENHANCEMENTS

Conclusion

LawBot effectively addresses the critical challenge of legal in-accessibility faced by millions of citizens across India by pro-viding a real-time, multilingual legal guidance interface. The system analyses user queries—whether typed or spoken—and retrieves contextually appropriate legal responses from a struc-tured, offline-capable knowledge base using keyword overlap scoring. By operating across eight languages and supporting both text and voice interaction, LawBot lowers the barrier for users unfamiliar with English or formal legal vocabulary. The system delivers responses through both text and text-to-speech, ensuring accessibility for users with low literacy. All things con-sidered, the experiment has demonstrated that AI-assisted legal information delivery can be an effective strategy for enhancing civic awareness and reducing legal helplessness. By providing a continuous, accessible information platform, LawBot can be de-

ployed on the web or mobile devices to help citizens understand and assert their rights.

Future Enhancements

Future developments could greatly increase accuracy and prac-tical impact, even beyond the current keyword-matching ap-proach. Integrating a fine-tuned large language model (LLM) for open-domain legal question answering would enable han-dling of queries beyond the pre-seeded knowledge base. Support for more Indian regional languages—such as Odia, Punjabi, and Assamese—would broaden reach. The system can be further improved with user authentication, personalised query history,

and case-tracking features. Integration with official govern-ment legal databases and real-time court judgement feeds would keep the information current. Proactive push notifications for important legal deadlines and rights could be delivered via a companion mobile application. In general, increasing the sys-tem's intelligence and breadth will improve civic empowerment and contribute to a more equitable access to justice.

REFERENCES

1. Firebase / Google Firestore Documentation. Google LLC. https://firebase.google.com/docs/ firestore.
2. Web Speech API — MDN Web Docs. Mozilla De-veloper Network. https://developer.mozilla.org/en-US/docs/Web/API/Web_Speech_API.
3. SpeechSynthesis API — MDN Web Docs. Mozilla De-veloper Network. https://developer.mozilla.org/en-US/docs/Web/API/SpeechSynthesis.
4. Python Official Documentation. Python Software Founda-tion. https://docs.python.org/3/.
5. Tailwind CSS Documentation. Tailwind Labs. https://tailwindcss.com/docs.
6. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of NAACL-HLT 2019, 4171–4186. https://arxiv.org/abs/1810.04805.
7. Phan, T. et al. (2023). AI-Driven Legal Chatbot Using Transformer-Based Question Answering. IEEE Access, 11, 34201–34215.
8. Majdi, A. et al. (2020). Multilingual NLP Pipeline for Domain-Specific Question Answering with Keyword Ex-traction and Rule-Based Matching. International Journal of Advanced Computer Science and Applications, 11(8), 112–120.
9. Sahayadhas, A. et al. (2021). Speech-Enabled Assistive Sys-tems for Low-Literacy Users: A Review of Voice Interface Designs and TTS Engines. Journal of Ambient Intelligence and Humanized Computing, 12, 5901–5918.
10. Sharma, R., & Verma, P. (2022). Real-Time Multilingual Legal FAQ System Using Firestore and Cosine Similarity Scoring. Procedia Computer Science, 204, 678–686.
11. Chollet, F. (2017). Deep Learning with Python. Manning Publications.
12. GitHub Repositories and Developer Forums (for reference and community support). https://github.com.