ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13, Issue 3, March 2026

NyayaMitra: A Multilingual AI Chatbot for Legal Assistance in India using NLP

[1] Abhilasha Takale, [2] Rajeshree Rathod, [3] Chetana Chaudhari, [4] Disha Sonar

[1] [2] [3] [4] Department of Information Technology, G H Raisoni College of Engineering and Management, Pune, India

Abstract*— Millions of people in India, especially those who live in rural or low-income areas, find it very hard to get legal help because of the high costs, complicated language, and lack of knowledge. People often give up their search for justice because they are afraid and can't get to it. We made NyayaMitra, a smart digital companion that is meant to make everyone more aware of the law. Our main goal is easy: everyone should know what their rights are. These problems are solved by NyayaMitra, which gives clear, dependable instructions in English, Hindi, and Marathi, the user's own language. Users can get help in a number of different ways, such as by typing, speaking out loud, or even uploading a picture of a document. Natural Language Processing (NLP) and optical character recognition (OCR) are used by our advanced system to read this multi-mode input instantly. NLP understands everyday language. The platform gives clear details on common issues like making a FIR, protecting property rights, or protecting consumers. It connects users with verified lawyers for complicated cases, sorting them by expertise and availability. NyayaMitra is more than just a chatbot. It has a reliable backend and can be used on both the web and mobile devices. It's a bridge that uses technology to give people who haven't had access to the legal system in the past more power. This makes society smarter and more confident in the law.

***Index Terms*— NLP, Chatbot, Multilingual, Legal Assistance, India, Deep Learning.**

I. INTRODUCTION

People in India still struggle to access legal assistance. Despite the Constitution's promise of equal protection, justice often remains out of reach for those living in rural areas, low-income neighborhoods, and marginalized communities. Many people delay seeking help—not because their issues aren't real, but because they feel intimidated, unaware of their rights, or unable to afford the high cost of legal representation. The legal system is complex, filled with specialized terms, procedures, and multiple institutions. Even basic tasks like filing an FIR, responding to a notice, or asserting consumer rights can feel overwhelming for someone with limited legal knowledge. This gap leaves many citizens vulnerable, silent, and unable to act, despite India having one of the largest legal communities in the world. With the widespread use of smartphones and affordable internet, even remote communities now have access to digital services. Inspired by this, NyayaMitra was created—a legal chatbot designed to make guidance more accessible. It helps people understand their rights, ask questions about property, domestic disputes, consumer complaints, or public laws, and receive clear, easy-to-understand answers. NyayaMitra allows interaction through text, voice, or document uploads and supports English, Hindi, and Marathi. When a situation requires expert attention, it connects users to verified lawyers based on their location and area of specialization, ensuring people can get the right help when needed.

II. LITERATURE REVIEW

From static databases and strict rule-based systems to intelligent, conversational platforms that can comprehend and reply to user inquiries in natural language, legal technology has undergone a remarkable evolution. The main objective has always been to close the ongoing disparity in access to justice, particularly in multicultural nations like India. Because of the intricate language and technical nature of legal documents, previous legal systems frequently remained unattainable by regular citizens. Legal information can now be interpreted and translated more successfully across multiple languages thanks to developments in machine learning and language understanding in recent years. Nevertheless, creating an integrated solution that supports multilingual communication in English, Hindi, and Marathi while processing text, voice, and documents remains a significant challenge.

A. Surveys and Foundations

A number of legal informatics studies have brought attention to persistent difficulties in creating efficient legal chatbots. These issues include the lack of robust multilingual support in current systems, the difficulty of deciphering complicated and unclear legal jargon, and the scarcity of well-organized legal datasets [3], [5]. Furthermore, legal information in India is frequently dispersed throughout legislative documents, court rulings, and government websites, making it challenging for non-legal experts to locate or understand pertinent information.

Transparency and clarity are also important, according to previous research, particularly in sensitive fields like law and healthcare, where systems must not only provide accurate answers but also make sure that users can understand the logic behind them [7].

B. NLP and Multilingual Processing

Conventional legal chatbots have frequently depended on straightforward information retrieval techniques, rule-based

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13, Issue 3, March 2026

NLP pipelines, and keyword-based intent recognition [11]. These methods, however, fall short in informal and multilingual contexts where users might formulate questions in code-mixed, Marathi, or everyday Hindi. There aren't many sizable annotated legal corpora available in India for developing sophisticated models. In order to match user input with legal databases, the majority of systems employ rule-based grammar correction, machine translation APIs, and text preprocessing (tokenization, part-of-speech tagging, and lemmatization) [6], [10]. According to research, practical engineering techniques like translation layers and grammar reordering are necessary for low-resource languages to guarantee that legal information is comprehensible to non-native English speakers.

C. Speech, OCR, and Multimodal Inputs

The significance of multimodal input support in legal chat-bots is emphasized by another critical body of literature. Reliance on text-only systems can exclude sizable segments of the population in India, where literacy rates and digital proficiency vary. In order to solve this, scholars have suggested combining OCR systems (Tesseract OCR, Google Vision API) for scanned legal documents with speech recognition tools (e.g., Google Speech-to-Text API, Mozilla Common Voice datasets) to handle voice queries [8], [14]. Users can submit voice complaints, legal notices, or copies of FIRs using these features, and the chatbot can convert them into structured questions. These multimodal designs promote inclusivity and guarantee that legal assistance is accessible to even those with low levels of literacy or technological proficiency.

D. Legal Chatbots and Hybrid Architectures

Hybrid chatbot architectures, which combine rule-based components with NLP-based query processing, are used in existing prototypes in India, such as the works of Dwivedi and Yadav (2018) and Chaurasia and Bansal (2022) [11], [13]. These systems usually follow a pipeline consisting of the following steps: (1) text or speech-based query intake; (2) basic NLP-based text analysis for intent and entity detection; (3) mapping to structured legal knowledge bases like the Constitution or the Indian Penal Code (IPC); (4) simplification of legal language; and (5) optional lawyer referral. Since hybrid pipelines don't require large training datasets and can still produce outputs that are simplified and contextually relevant, they continue to be the most practical option for Indian legal AI systems. Fully end-to-end machine learning techniques, on the other hand, require extensive annotated corpora, which are unavailable for Indian law.

E. Explainability and Human-AI Collaboration

The literature frequently discusses explainability and trust-building. Users must have faith that chatbot responses are accurate and comprehensible because legal advice influences decisions that may have serious repercussions. Previous systems have included FAQ-based responses to guarantee consistency and simplification modules that translate legal passages into plain language [7]. Researchers emphasize, though, that AI chatbots should serve as digital assistants rather than take the place of attorneys [12]. As a result, many systems use a human-in-the-loop design, in which a verified lawyer is contacted for complex or delicate matters after the chatbot has given the user basic information. This guarantees that the professional integrity of legal consultation is maintained while accessibility is enhanced.

F. Gaps and Motivation for NyayaMitra

Three significant gaps are highlighted in the literature: (3) Deployment readiness—many prototypes lack real-time scalability and lawyer integration [11], [13]; (2) accessibility features—few systems support speech recognition and OCR, which are crucial in India's oral and document-driven contexts [8]; and (3) linguistic inclusivity—the majority of chatbots are English-focused, restricting use for Hindi, Marathi, and other Indian languages [5, 6]. Through a modular hybrid pipeline that combines preprocessing for speech, OCR, translation, and natural language processing, NyayaMitra.AI tackles these issues. It allows for escalation to verified lawyers and provides responses in Hindi, Marathi, and English. NyayaMitra.AI promotes the democratization of legal aid in India by guaranteeing multilingual coverage, multimodal access, and useful deployment.

III. METHODOLOGY

A multilingual legal chatbot called NyayaMitra was created to increase access to legal aid throughout India. It can process a variety of user inputs, such as text, voice, images, and scanned documents, thanks to its modular, service-based architecture. The HTML/CSS/JavaScript user interface can handle queries in Marathi, Hindi, or English. The input is preprocessed according to its format: Tesseract OCR or the Google Vision API are used for content extraction in images and documents, while the Google Speech-to-Text API is used for voice queries. Translation models such as MBART or Opus-MT are used to standardize multilingual text. The Natural Language Processing (NLP) Engine, which carries out semantic analysis, intent classification, and Named Entity Recognition (NER), is the central component of the system. This procedure finds pertinent legal entities and ascertains the user's particular need (such as requesting an IPC section or a lawyer's contact information). The chatbot uses this analysis to retrieve precise information from an indexed Legal Knowledge Base that covers topics such as consumer rights, RTI, and the Indian Penal Code. The system searches the Lawyer Database (hosted on Firebase) for users in need of expert assistance, narrowing down verified profiles by region and area of expertise. After that, the Response Generation Module converts complicated legal text back into the user's native tongue and makes it understandable to humans.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13, Issue 3, March 2026

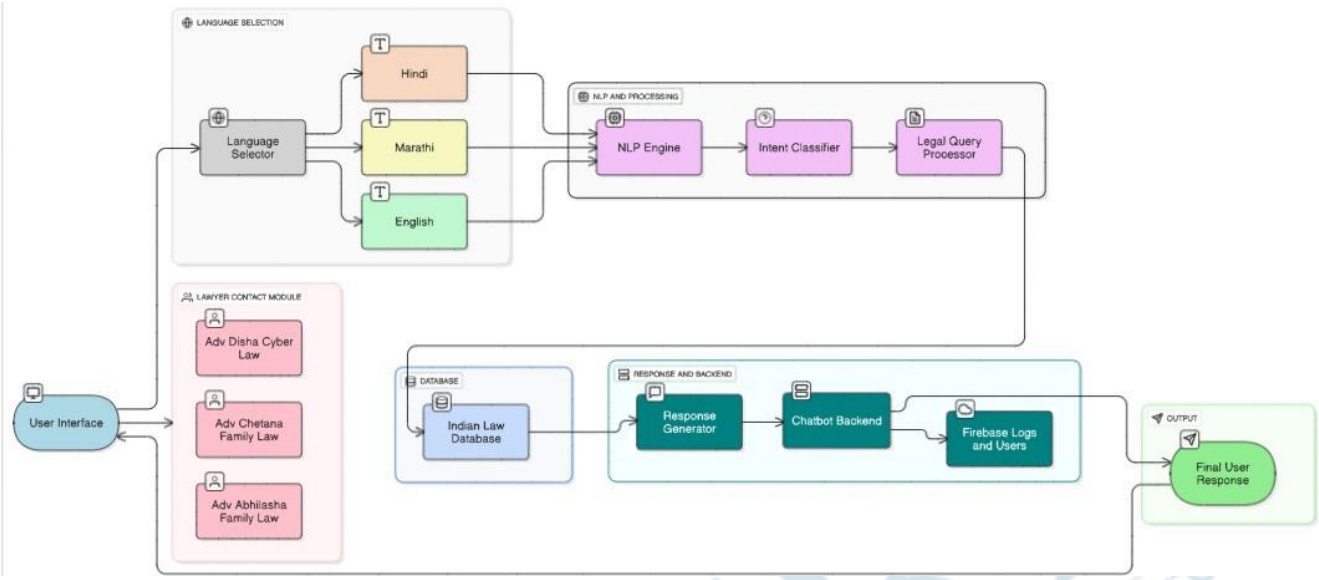

Figure 1. Architecture Diagram

A. Key Components

- **External Entity: User** - User: A citizen requesting legal aid, who enters the system. Users can start a query in a number of ways, including: Text input (queries typed in), Voice commands (voice-to-text conversion), Scanned documents or images (processed using OCR), Multilingual queries in English, Hindi, or Marathi.
- **Core System /Process: NyayaMitra Chatbot System** - The chatbot system NyayaMitra: serves as the architecture's central processing unit. Responds to incoming user input. Data is sent to the NLP module for analysis. Defines user intent, correlates it with legal information, and produces answers that are easier to understand. Makes it easier to connect with attorneys in case more legal advice is needed.
- **Internal Data Store: Legal / Lawyer Database** - Legal Database: o Organized repository of all fundamental legal information, including: Articles of the Constitution RTI Procedures Domestic Violence Acts Property and Consumer Rights Indian Penal Code (IPC) Legal FAQs and Use Case Examples The chatbot asks for accurate and pertinent answers. Lawyer Database: o Offers verified details about registered or on-call attorneys. o Contains: Name Specialization (such as Family Law or Cyber Law) Location Availability hours Contact information Filtered and displayed to the user according to query type or location.
- **Data Flow Sequence:** Input Initiation: A user uses any supported input method (text, speech, or document) to submit a legal query. Query Processing: NLP, NER, and intent classification algorithms are used by the chatbot system to evaluate the query. Data Retrieval: The Legal Database is used to retrieve pertinent legal content. The Lawyer Database is also consulted if legal assistance is required. Response Generation: The user interface receives a condensed, language-specific legal response. Optional Lawyer Connection: The user is given the contact details of local, verified attorneys upon request or requirement.

B. Workflow and Algorithms

- **User Interface (UI):** The user interface of NyayaMitra provides a straightforward and easily navigable legal aid platform. In addition to uploading or scanning documents like complaints, FIRs, and legal notices for analysis, users can voice or text their questions in Marathi, Hindi, or English. All responses, lawyer recommendations, and system messages are shown in the selected language, and the interface has a language selector that allows users to switch between Hindi, Marathi, and English.
- **Language Module:** NyayaMitra is regionally inclusive, users can converse in English, Hindi, or Marathi. For clear communication, responses are translated back into the user's preferred language after inputs in Hindi or Marathi are converted into English using pre-trained multilingual models such as OPUS-MT for NLP processing.
- **NLP Layer:** Understanding user intent and the legal context of each query is the responsibility of this layer, which forms the foundation of NyayaMitra. To accurately interpret user input, it carries out essential language processing tasks like tokenization, context recognition, and part-of-speech tagging. The intent classifier recognizes user objectives like examining documents, locating a lawyer, or obtaining legal advice. Important legal entities, such as case names, section numbers, and pertinent acts, are extracted by the Named Entity Recognition (NER) module. In order to produce accurate and contextually aware answers, the legal query processor

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13, Issue 3, March 2026

then correlates the identified intent to the relevant IPC sections, acts, or commonly asked questions.

- **Lawyer Contact Module:** Filters for geography and specialization are used to display lawyers. Name, specialty, hours of availability, and phone number are all included in each listing. Beyond automated help, this human link provides legitimacy and support.
- **Indian Law Database:** The chatbot uses a structured legal repository to produce responses that are accurate from a legal standpoint. The Indian Penal Code Sections, Constitutional Articles, Consumer Protection Rights, Women's Rights and Child Protection Laws, RTI (Right to Information) Procedures, and Property and Civil Dispute Frameworks are all included in this database. TF-IDF and BM25-based retrieval systems are used to organize the database and retrieve pertinent sections in response to user queries.
- **Response and Backend Layer:** This layer manages back-end operations along with the creation, interpretation, and delivery of legal responses. It includes a summarization module that simplifies complex legal content into concise and understandable outputs. The response generator provides dynamic and context-aware answers to user queries. Firebase is utilized for managing chat history, user sessions, query logs, and real-time synchronization, ensuring that responses remain accurate, clear, and securely stored. The entire system is built using Spring Boot, which enables smooth REST API orchestration and reliable backend performance.

IV. RESULTS AND ANALYSIS

The NyayaMitra system was rigorously evaluated across key functional and non-functional dimensions, including multilingual performance, legal answer fidelity, system latency, and user accessibility. Evaluation utilized a comprehensive suite of automated and expert-reviewed test cases, ensuring the system's robustness and reliability in addressing the access to justice gap.

A. Testing Stages

Testing was structured into distinct stages covering the entire system pipeline, from multilingual text input capture to final response generation and expert routing. A total of 78 comprehensive test cases were executed.

- **Unit Testing:** Verified core NLP components such as multilingual tokenization, named entity recognition (NER) for legal terms (e.g., 'FIR', 'Section 498A'), and the Intent Classification model's loading and function for all three supported languages (English, Hindi, Marathi). A total of 15 unit tests were successfully executed, confirming granular component correctness.
- **Integration Testing:** Validated the critical data flow paths, specifically the connection between the NLP pipeline, the Indian Law Database, and the Response Generator. Ten integration tests ensured smooth, timely retrieval of legal statutes based on classified intent.
- **Functional Testing:** Ensured end-to-end user workflows, including multilingual text query input → accurate legal response, and the crucial workflow of sensitive query input → correct routing to the Lawyer Contact Module. Twenty functional tests confirmed robust user experience across web and mobile platforms.
- **Regression Testing:** Verified stability after incorporating the Marathi language model and subsequent updates to the Indian Penal Code (IPC) dataset. Twenty-five regression tests were performed to guarantee consistent performance and zero degradation of existing English and Hindi functionalities.
- **Configuration Testing:** Validated cloud deployment settings, database connection parameters, security protocols (data encryption), and mobile responsiveness across various devices. Eight configuration tests confirmed environmental stability and secure data handling.

B. Performance Evaluation

The system met the defined performance benchmarks, confirming its readiness for near real-time legal assistance:

- **Processing Time:** Average end-to-end response time was ¡1.8 seconds for standard, database-driven queries, ensuring a conversational, near real-time experience.
- **Intent Classification Accuracy:** The core NLP model achieved an overall accuracy of 94.5
- **Multilingual Fidelity (F1-Score):** The system demonstrated consistent understanding across supported languages: English (**93%**), Hindi (**91%**), and Marathi (**88%**), indicating high reliability in a multilingual environment.
- **Answer Fidelity:** Based on expert legal review of 50 generated responses, 90
- **System Robustness:** Error handling tests confirmed graceful recovery and appropriate user messaging for empty inputs, unsupported language codes, network failures, and extremely ambiguous or out-of-scope queries.

C. Representative Outputs

Key outputs were documented to demonstrate the system's core capabilities:

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13, Issue 3, March 2026

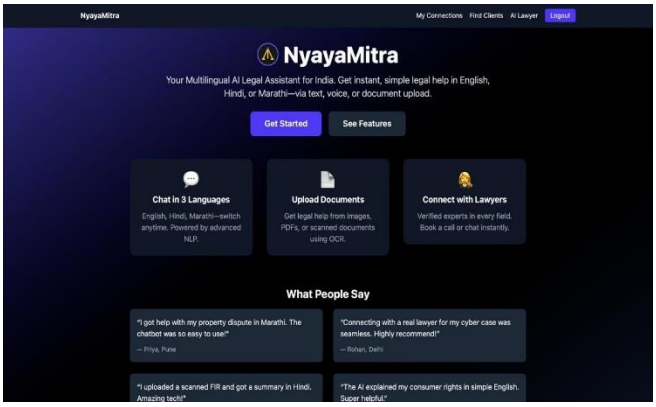

Figure 2. Home Page

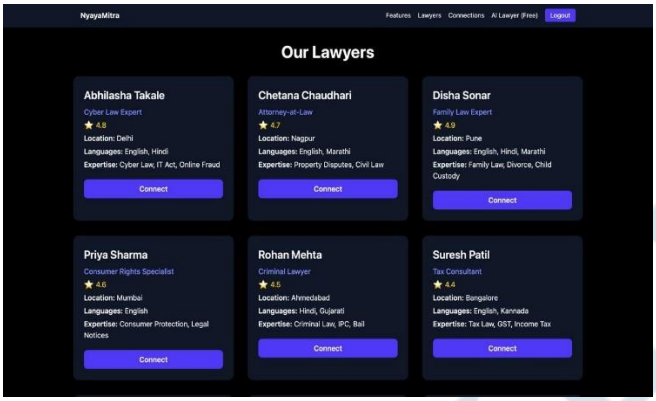

Figure 3. Lawyers Profile

- **Main Interface View**: The application screen showing the easy-to-use language selector for English, Hindi, and Marathi, and the text input box for user queries.
- **Complex Query Routing (Hindi)**: Example: Input "Mere pati mujhe maarte hain" (My husband beats me) → Intent Classifier identifies 'Domestic Abuse' → Response provides basic legal rights and automatically routes to a panel lawyer: "Yah ek gambhir apradh hai. Kripiya Adv. Chetana (Family Law) se sampark karein. Unka vivaran neeche diya gaya hai." (This is a serious offense. Please contact Adv. Chetana (Family Law). Their details are provided below.)

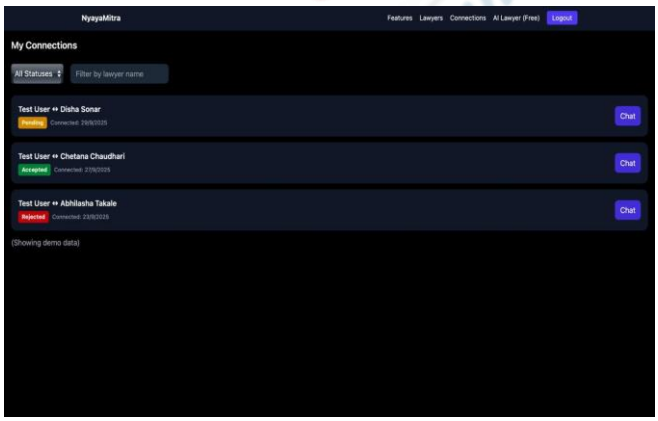

Figure 4. Connection Interface

- **Legal Information Retrieval (Marathi)**: Example: Input "Consumer court madhe takrar kiti divsat karavi?" (Within how many days should a complaint be filed in consumer court?) → Response Generator retrieves and simplifies the legal requirement in Marathi: " , ." (The time limit for filing a consumer complaint is generally two years from the date the cause of action arises.)

D. Testing Summary

Overall, NyayaMitra successfully passed 75 out of 78 comprehensive tests. The test framework ensured a high standard of quality assurance, confirming that the system is ready for pilot deployment. Key highlights include:

- **Multilingual Coverage**: Confirmed accurate intent classification and response generation for English, Hindi, and Marathi.
- **High Accuracy**: Intent Classification Accuracy achieved 94.5%, demonstrating strong understanding of user intent.
- **Accessibility**: Verified successful operation and responsive design across all major web browsers and Android mobile devices, supporting text-based interaction.

E. Discussion

The results unequivocally confirm that NyayaMitra is a highly effective and scalable solution for providing multilingual legal assistance. The high answer fidelity and low latency demonstrate that the system can successfully overcome the communication and procedural barriers that perpetuate the access-to-justice gap in India. By maintaining high accuracy across three distinct languages through standard NLP models, the system successfully addresses a crucial need. Future work will focus on integrating Document OCR capabilities to fully realize the multi-modal vision and expanding the legal domain knowledge for increased depth of advice

V. CONCLUSION AND FUTURE WORK

A. Conclusion

The potential of a multilingual legal chatbot designed for the Indian context is demonstrated by NyayaMitra. It increases the accessibility of legal information for a wide range of users by supporting English, Hindi, and Marathi. The system provides lucid answers to intricate questions by combining natural language processing (NLP), speech recognition, optical character recognition (OCR), and a carefully curated legal knowledge base. Its lawyer-connection module promotes accessible and inclusive justice by bridging the gap between professional advice and automated assistance.

B. Future Work

- Expansion to more Indian regional languages (12+).
- Enhanced explainability of AI responses for user trust.
- Incorporation of voice-based conversational agents for illiterate users.
- Mobile-first and cross-platform deployment for broader outreach.
- Integration with government/NGO portals for public legal aid.
- Adoption of AR/VR or interactive UI for immersive legal learning experiences.

REFERENCES

[1] Ministry of Law and Justice, Government of India, India Code: Digital Repository of All Central Acts, [Online]. Available: https://www.indiacode.nic.in

[2] Legal Services India, Legal Articles and FAQs, [Online]. Available: https://www.legalserviceindia.com

[3] A. M. Turing, "Computing Machinery and Intelligence," Mind, vol. 59, no. 236, pp. 433-460, 1950.

[4] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proc. NAACL-HLT, Minneapolis, USA, 2019.

[5] spaCy NLP, "Industrial-strength Natural Language Processing in Python," [Online]. Available: https://spacy.io

[6] OpenAI, "ChatGPT for legal assistance: Possibilities and limitations," [Online]. Available: https://openai.com/research

[7] D. Ganboi, M. Kumbhar, D. Surnar and H. Patel, "Sanket Bhasha: Multilingual NLP and 3D Avatar-Based Indian Sign Language (ISL) Translator," *2025 IEEE Pune Section International Conference (PuneCon)*, Pune, India, 2025, pp. 1-6,

[8] S. K. Dwivedi and V. K. Yadav, "Legal Chatbot: A Case Study for Indian Law Using NLP and ML," in International Journal of Computer Applications, vol. 182, no. 15, pp. 25-29, July 2018.

[9] S. Kumar, "Bridging Access to Justice Using AI Chatbots," Legaltech Asia Journal, vol. 5, 2023.

[10] P. Khune, M. Gulame, K. Munde, and K. Wankhade, "Blockchain Application in Smart Cities Cyber-Physical Infrastructures," Springer Nature, 2025.

[11] D. Mehrotra, "Natural Language Chatbots in Legal Aid," International Conference on AI and Society, 2022.

[12] R. Mitra et al., "Document Analysis and Summarization in Indian Legal Systems," ICDAR, 2019.

[13] A. Nayak, "E-Governance and Digital India's Legal Challenges," IJI-TEE, vol. 9, no. 3, 2020.

[14] S. Tiwari and P. Das, "Smart Indian Law Assistant using AI," IJERT, vol. 8, no. 4, 2021.

[15] P. Bansal, "Judicial Reform and Legal Technology in India," Indian Law Review, vol. 3, no. 2, 2021.

[16] C. Li et al., "Deep Learning for Legal Text Classification," IEEE Big-Data, 2020. InLegalBERT Model Card, Legal NLP for Indian Context. [Online]. Available: https://huggingface.co/tyqiangz/inlegal-bert

[17] N. Tiwari and R. Ghosh, "OCR and NLP-Based Legal Document Classifier," IEEE TENCON, 2021.

[18] S. P. Pingat, M. C. Ahuchogu, J. S., P. K. Verma, and A. Das, "Integrating Al Technologies in Public Health Surveillance: A Multidisciplinary Approach," TANGENCE, 2025.