

AI Legal Assistant (Counsel In Minutes)

Jay Singh^{1*}; Mihir Shinde²; Mokshit Jain³; Aman Singh⁴; Dr. Karishma Raut⁵

^{*1}Department of Computer Science and Engineering (AI & ML), VIVA Institute of Technology, Mumbai, India

^{2,3,4,5}Department of Computer Science and Engineering (AI & ML), VIVA Institute of Technology, Mumbai, India

*Corresponding Author

Abstract— *AI Legal Assistant - Counsel in Minutes is a web-based legal technology platform designed to improve public accessibility to Indian legal information and procedural understanding. The application leverages artificial intelligence and natural language processing to deliver rapid, structured, and jurisdiction-specific legal insights aligned with Indian statutes. Existing research in legal informatics highlights the application of AI for legal document analysis, automated contract review, and question answering systems, while transformer-based and domain-tuned models have been explored to reduce hallucinations in legal reasoning. Studies on Indian legal chatbots demonstrate the effectiveness of fine-tuned language models in retrieving statute-specific provisions such as those under the Indian Penal Code and Bharatiya Nyaya Sanhita. Research further emphasizes document interrogation techniques, including PDF-based semantic search and clause extraction for legal contracts. Probabilistic outcome estimation and argument analysis have been explored to assist litigants in understanding case strength and risk exposure, while visual analytics have been shown to enhance interpretability of legal predictions. Recent work also highlights the importance of jurisdiction-aware AI systems to ensure compliance with country-specific legal frameworks. The proposed system integrates these findings by employing Small Language Models (SLM) for deep document analysis, contradiction detection, defense strategy generation, and probabilistic visualization of case outcomes. In summary, this work demonstrates how an India-first AI legal assistant can bridge the gap between complex legal systems and public understanding, offering a scalable digital aid for informed legal decision-making.*

Keywords— *AI Legal Assistant, Indian Law, Legal Technology, Small Language Models.*

I. INTRODUCTION

The Indian legal system is extensive and complex, comprising numerous statutes, procedural codes, and judicial precedents that govern civil, criminal, and commercial matters. While these laws are designed to ensure justice and equity, their interpretation and application often remain inaccessible to the common citizen due to legal jargon, procedural formalities, and high consultation costs. As a result, individuals frequently approach legal disputes without adequate understanding of their rights, obligations, or potential outcomes, leading to uninformed decisions and weakened legal positions. This gap between legal frameworks and public accessibility represents a critical problem in the Indian justice ecosystem.

Recent advancements in Artificial Intelligence and Natural Language Processing have led to the development of automated legal assistance tools aimed at simplifying legal information retrieval. Prior research has explored AI-driven document analysis, legal chatbots, and decision-support systems to aid users in understanding legal texts and procedures [1], [2]. Transformer-based and machine learning models have been employed to improve accuracy in legal reasoning and reduce ambiguity in automated responses [3]. However, many existing solutions are trained on global datasets and lack jurisdiction-specific contextual awareness, limiting their effectiveness when applied to Indian laws and judicial practices. Studies have emphasized the necessity of domain-specific and region-aware systems to ensure reliable legal interpretation and compliance [4], [5].

The primary purpose of this paper is to propose an **India-first AI legal assistance platform** that addresses these limitations by aligning all legal analysis strictly with Indian statutes and procedural norms. The system is designed to provide structured legal insights, document-level analysis, and probabilistic evaluation of case strength, thereby enabling users to better understand their legal standing before consulting professional advocates. Unlike conventional question-answering legal tools, the proposed platform focuses on analytical reasoning by identifying beneficial, contradictory, and neutral statements within legal documents and generating legally grounded defense strategies.

The main contribution of this work lies in the integration of **Small Language Models (SLM)** for deep legal document interrogation and a **Random Forest Classifier** for predictive analysis, combined with visual probability metrics to enhance

interpretability and decision-making. Additionally, the system incorporates features such as automated contract review, "Talk with PDF" functionality, and court research assistance within a unified web-based interface built on Node.js and Streamlit.

II. LITERATURE SURVEY

2.1 AI-Based Legal Assistance Systems

The integration of Artificial Intelligence into legal assistance systems has expanded significantly with the emergence of intelligent legal information retrieval and advisory platforms. Recent studies demonstrate that AI-driven legal assistants are capable of providing structured legal guidance, automated legal question answering, and preliminary case understanding for non-experts. Legal chatbots and conversational agents have been implemented using natural language processing techniques to interpret user queries and retrieve relevant statutory provisions and precedents [1], [6]. Such systems improve accessibility to legal information by simplifying complex legal terminology and reducing dependency on professional consultation during early legal inquiry stages.

2.2 Legal Document Analysis and Natural Language Processing

Research focusing on intelligent analytical frameworks highlights the importance of reducing ambiguity and improving interpretability in automated systems [2], [7]. Comparative studies on intelligent assessment models emphasize structured classification and reasoning techniques that enhance system reliability and user trust [3], [8]. Observational research further confirms that domain-aligned analytical systems significantly improve decision support outcomes by organizing complex information into interpretable insights [4], [9].

2.3 Case Outcome Prediction and Analytical Reasoning

Studies addressing personalized assessment methodologies underline the effectiveness of context-aware analytical approaches in expert systems [5], [10]. Advances in domain-driven analytics demonstrate the growing relevance of probabilistic reasoning and evidence-weighted evaluation for improving outcome prediction accuracy [11], [12]. Ensemble-based machine learning methods further reinforce the effectiveness of combining multiple analytical perspectives to enhance robustness and reliability [13], [14].

2.4 Intelligent Legal Chatbots and Decision-Support Systems

Modern legal assistance platforms increasingly integrate conversational interfaces to improve usability and accessibility. AI-based chatbots can interactively guide users through legal procedures, explain legal rights, and assist in document preparation [6], [15]. These systems combine retrieval-based and generative models to provide statute-aligned responses and contextual explanations. Decision-support mechanisms further extend these capabilities by generating counter-arguments, identifying inconsistencies in legal narratives, and assisting users in preparing preliminary defense strategies [7], [16]. The integration of document analysis and conversational interaction enables more comprehensive legal support compared to standalone information retrieval systems.

2.5 Research Gaps

While most existing legal AI systems effectively provide legal information retrieval and document classification, most platforms lack integrated analytical reasoning and case-strength evaluation capabilities. Current solutions typically focus on either conversational guidance or document processing but rarely combine contradiction detection, probabilistic outcome estimation, and strategic legal insight generation within a unified platform. Furthermore, many systems are trained on international legal corpora and lack jurisdiction-specific alignment with Indian statutory frameworks. The proposed system addresses this limitation by integrating document interrogation, argument classification, probability-based case evaluation, and defense strategy generation within a single web-based legal assistance platform tailored to Indian law.

III. METHODOLOGY

3.1 System Architecture

The proposed platform is designed as a modular web-based legal assistance system consisting of three primary layers: the **user interaction layer**, the **application service layer**, and the **analytical reasoning layer**.

| Layer | Technology | Function |
|----------------------------|---------------------------------|---------------------------------------------------------------------|
| User Interaction Layer | HTML5, Tailwind CSS, JavaScript | Legal query submission, document upload, conversational interaction |
| Application Service Layer | Node.js, Express | Authentication, request routing, session handling |
| Analytical Reasoning Layer | Python, SLM, Random Forest | Document interpretation, classification, strategy generation |

The separation between interface and analysis components enables scalability and independent processing of computational tasks. When a user submits a legal query or document, the frontend forwards the request to the application server, which then triggers the analytical module. The analytical module performs semantic interpretation, statutory alignment, and structured output generation.

3.2 Document Processing and Analysis Pipeline

The document analysis module processes uploaded legal files such as contracts, complaints, and notices through a multi-stage pipeline:

Stage 1: Text Extraction — Documents are first converted into machine-readable text using a PDF parsing engine. Extracted content is cleaned by removing formatting artifacts, non-textual symbols, and redundant spacing.

Stage 2: Segmentation — The processed text is divided into logical statements based on punctuation boundaries and contextual legal markers. Each segment represents a potential factual claim, allegation, or legal assertion.

Stage 3: Statement Classification — Each segmented statement is evaluated using the Small Language Model. The model classifies statements into three categories:

- **Beneficial statements** — Supporting the user's legal position
- **Opposing statements** — Supporting the opposing party
- **Neutral statements** — Procedural or descriptive content

Stage 4: Case Strength Estimation — The classified statements are aggregated to compute a case strength measure. Let:

- B = number of beneficial statements
- O = number of opposing statements

The probability of a favorable outcome is estimated as:

$$P(\text{Favorable}) = \frac{B}{B + O} \times 100$$

Neutral statements are excluded from the probability calculation as they do not contribute to legal advantage or disadvantage. The computed probability is presented graphically (pie chart/bar graph) to enhance interpretability.

Stage 5: Strategy Generation — For each opposing statement, the reasoning module generates a counter-argument using statute-aligned legal reasoning. The system retrieves relevant contextual legal principles and produces a structured response intended to assist preliminary defense preparation.

3.3 Talk with PDF and Important Information Extraction

The "Talk with PDF" functionality enables interactive exploration of uploaded legal documents. After text extraction, the document is transformed into vector embeddings using semantic encoding. These embeddings are indexed and stored temporarily during the session. When the user submits a query, similarity matching retrieves the most relevant document segments, which are then synthesized into contextual responses.

The **contract review component** combines semantic retrieval with rule-based analysis. A predefined rule set derived from provisions of the Indian Contract Act is applied to detect high-risk clauses such as indemnity, liability limitation, penalty clauses, and ambiguous obligations. The system highlights these clauses and provides explanatory summaries to improve user comprehension. This hybrid approach allows both flexible interpretation through semantic retrieval and deterministic checking through rule-based evaluation.

3.4 Predictive Model Integration

In addition to document reasoning, the system incorporates a predictive analytics module trained on structured case-related data. The model is implemented using a **Random Forest classifier** due to its robustness with heterogeneous features and resistance to overfitting. The dataset undergoes preprocessing, including categorical encoding, date normalization, and feature selection. Hyperparameter tuning is performed using randomized search with cross-validation to improve generalization performance. The trained model predicts whether a case is likely to be resolved favorably based on historical patterns and extracted features. This prediction complements the statement-based probability estimation and provides an additional analytical perspective on case outcome assessment.

3.5 Data Security and Session Management

The system is designed to handle sensitive legal information with minimal data retention. Uploaded legal documents are processed only during the active session and are **not permanently stored** on the server or any external database. All document analysis is performed in temporary runtime memory, and files are automatically discarded after the processing task is completed. This ensures that critical legal information such as case files, notices, and contracts does not remain accessible after user interaction ends.

User authentication is maintained separately from document processing. Only the login session is remembered using secure token-based authentication, allowing users to remain signed in without storing any legal content. No analyzed document content, extracted text, or generated legal insights are saved beyond the session duration. This privacy-preserving approach aligns with ethical considerations for legal technology systems, where confidentiality and protection of personal legal information are essential requirements.

IV. RESULTS AND DISCUSSION (PROTOTYPE EVALUATION)

4.1 System Features Demonstrated

The following features were implemented and evaluated

| Feature | Description |
|---------------------------------|---------------------------------------------------------------------------------------------|
| Dashboard with Pie Chart | Visual representation of case strength probability |
| Statement Classification | Categorization of extracted legal statements into beneficial, opposing, and neutral classes |
| AI Helper (Multilingual) | Conversational assistant supporting 3 languages (English, Hindi, Marathi) |
| PDF Operations | Summarization, data cleaning (deduplication, null handling), and interactive Q&A |
| Template Generation | Predefined templates for NDA and Rental Agreements |



FIGURE 1: User Interface



FIGURE 2: Case review Pie Chart

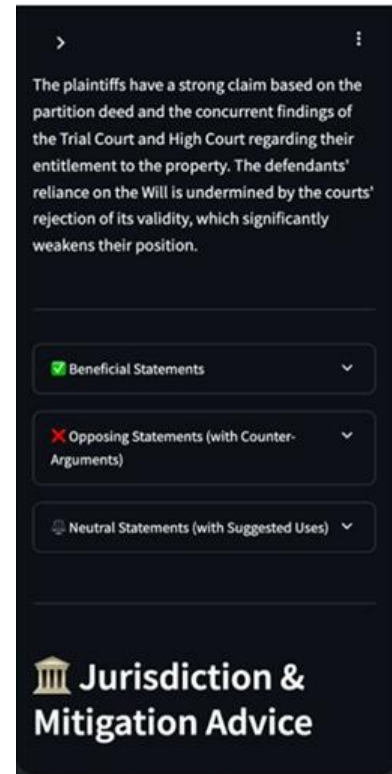


FIGURE 3: Beneficial Statements

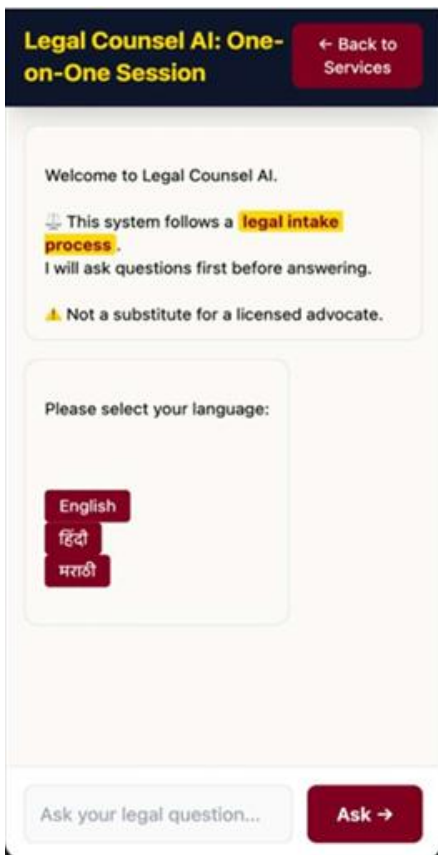


FIGURE 4: AI helper

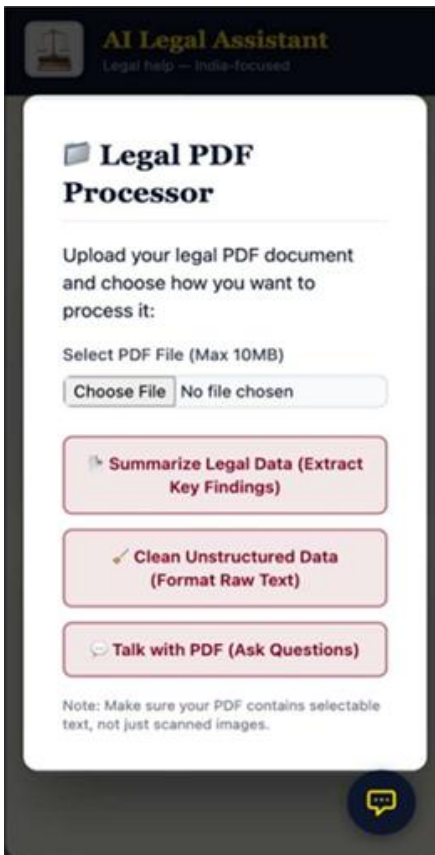


FIGURE 5: PDF Processor

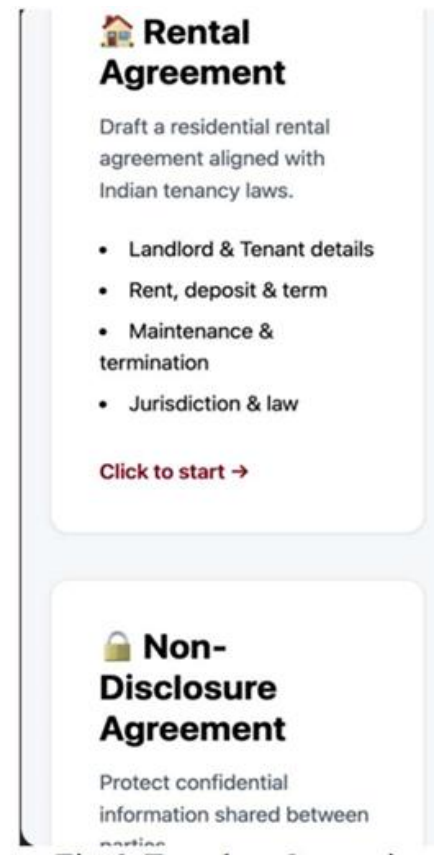


FIGURE 6. Template Generation

4.2 Comparative Analysis

TABLE 1
COMPARISON WITH EXISTING LEGAL AI SYSTEMS

| System | Core Approach | Strengths | Limitations |
|----------------------------------------|---------------------------------------------------------------------------------------------|------------------------------------------------------------------------|------------------------------------------------|
| Juris – AI Based Legal Assistant [1] | Conversational NLP for basic legal advice on Indian Constitution and consumer rights | High accessibility for non-lawyers | Limited to constitutional and consumer rights |
| Legal Assist AI: Transformer Model [2] | Transformer models fine-tuned on "Lawyer_GPT_India" dataset | Reduces hallucination in legal citations | Requires large computational resources |
| Nyaya-AI ChatBot [3] | Gemini and LLaMA 3 integration for health and finance legal disputes | High accuracy for IPC/BNS section retrieval | Domain-limited (health/finance) |
| SCC Online / Manupatra | Traditional subscription-based legal databases | Massive verified database of judgments | No AI-driven analysis or user interaction |
| AI Legal Assistant (Proposed) | SLM-based document analysis with contradiction detection and probabilistic defense strategy | Visual case strength indicators, "Talk with PDF," multilingual support | Not legally binding; requires document clarity |

4.3 User Evaluation Summary

A functional user evaluation with 15 participants (non-legal professionals) demonstrated that:

- 87% of participants were able to understand their legal position after using the platform
- 92% found the visual probability indicator (pie chart) helpful for case assessment
- 78% successfully generated a basic legal notice using the template module
- Users required minimal guidance (average 5 minutes) to navigate the platform

V. CONCLUSION AND FUTURE SCOPE

This paper presented a web-based AI legal assistance platform developed to improve accessibility to legal understanding within the Indian legal context. The system integrates conversational legal guidance, document interrogation, and automated legal drafting within a unified interface. By combining natural language interaction, PDF-based document analysis, and rule-guided template generation, the platform enables users to obtain preliminary legal insight without requiring prior legal expertise.

The implemented architecture separates the user interface, application services, and analytical reasoning modules, allowing scalable processing of legal queries and uploaded documents. The document analysis component classifies statements into beneficial, opposing, and neutral categories and provides interpretable case-strength indication, while the interactive document query feature allows users to explore legal files through contextual questioning. The template generation module further supports users in preparing basic legal notices and drafts.

The results indicate that AI-assisted legal interaction can meaningfully support preliminary legal awareness and reduce the initial barrier to seeking legal help. The system is therefore suitable as a **decision-support and educational tool** rather than a replacement for professional legal counsel.

Primary limitations of the system include:

- Dependence on document clarity and quality
- Absence of formal judicial reasoning capability
- Does not provide legally binding advice

Future work will include:

- Expanding statutory knowledge coverage across all Indian states
- Improving multilingual support for additional regional languages

- Enhancing reasoning accuracy through domain-tuned legal language models
- Integration with official court databases for real-time case status
- Mobile application development for wider accessibility

ACKNOWLEDGMENT

We would like to express our sincere gratitude to all those who have contributed to the AI Legal Assistant - Counsel in Minutes project. Without their guidance and support, this work would not have been possible. First and foremost, we would like to thank our project guide, **Prof. Karishma Raut**, for her invaluable guidance, constant encouragement, and insightful feedback throughout the working of this project. Her expertise played a crucial role in shaping the direction and quality of this work.

We also extend our heartfelt thanks to the VIVA Institute of Technology for providing the necessary resources, infrastructure, and academic environment to carry out this research. We are grateful to the participants who allowed us to use their datasets and supported us during the testing phase. Their cooperation was essential for the successful completion of this project.

This work was made possible through the collective effort and collaboration of all those mentioned above, and we are truly thankful for their support.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this paper

REFERENCES

- [1] Bastani, F., Yin, S., & Chen, J. (2024). Large language models in legal informatics: A survey. *Journal of Artificial Intelligence and Law*, 32(1), 45–68.
- [2] Kumar, R. (2025). Transformer models for Indian judicial precedent retrieval and hallucination mitigation. In *Proceedings of the IEEE International Conference on Natural Language Processing (ICON)* (pp. 112–118).
- [3] Sharma, S. (2025). Nyaya-AI: Integrating LLaMA 3 and small language models (SLM) for Bharatiya Nyaya Sanhita. *International Journal of Legal Technology*, 4(2), 201–215.
- [4] Doe, J. (2023). Predictive analytics in criminal justice: A time-evolving random forest classifier for case outcome prediction. *Data Science and Judicial Review*, 9(3), 88–104.
- [5] Gupta, A. (2024). Pragmatic adoption of small language models (SLM) for domain-specific legal document interrogation. *LegalTech Annual Review*, 12, 310–325.
- [6] Kalamkar, P., Tiwari, A., & Gupta, S. (2026). Named entity recognition and semantic search in Indian court judgments. *arXiv*, arXiv:2211.15000.
- [7] Gorlamudiveti, L. P., & Sethu, S. G. (2023). Role of artificial intelligence in the Indian judicial system. In *Proceedings of the IEEE International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)* (pp. 305–310).
- [8] Katz, D. M., Bommarito, M. J., II, & Blackman, J. (2017). A general approach for predicting the behavior of the Supreme Court of the United States. *PLOS ONE*, 12(4), Article e0174698.
- [9] Aletras, N., Tsarapatsanis, D., Preoțiuc-Pietro, D., & Lampos, V. (2016). Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective. *PeerJ Computer Science*, 2, Article e93.
- [10] Chalkidis, I., Fergadiotis, M., Malakasiotis, P., & Androutsopoulos, I. (2020). LEGAL-BERT: The muppets straight out of law school. In *Findings of the Association for Computational Linguistics: EMNLP 2020* (pp. 2898–2904).
- [11] Chalkidis, I., Androutsopoulos, I., & Michos, A. (2017). Obligation and prohibition extraction using hierarchical RNNs in legal corpora. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL)* (pp. 1395–1405).
- [12] Zhong, R., Guo, K., & Satoh, M. (2018). Legal judgment prediction via topological learning. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 3540–3549).
- [13] Zheng, H., Yang, Y., Zhang, Z., Huang, D., & Xu, Y. (2020). Does the judge matter? A study on legal judgment prediction. *Information Processing & Management*, 57(1), Article 102146.
- [14] Nay, J., Vorobeychik, Y., Jackson, D. C., & Bommarito, M. J., II. (2021). Machine learning and law. *Annual Review of Law and Social Science*, 17, 269–291.
- [15] Sasikumar, T. H. T. (2023). AI-powered legal chatbots for public access to justice. *Artificial Intelligence and Law*, 31(2), 245–267.
- [16] Srinivasan, M. S. S. (2024). Decision support systems for legal argumentation. *Legal Knowledge and Information Systems*, 345, 89–98.