

NyaySetu: An AI-Assisted Legal Suggestion System

Anupam Kumar Singh
Computer Science & Engineering
Sharda University
Greater Noida, India

Akshay Pratap Singh
Computer Science & Engineering
Sharda University
Greater Noida, India

Prince Kumar
Computer Science & Engineering
Sharda University
Greater Noida, India

Dr. Velayudham Sathiyasuntharam
Professor
Computer Science & Engineering
Sharda University
Greater Noida, India

ABSTRACT - Access to legal assistance remains out of reach for many individuals due to high costs, limited awareness of legal rights, and the intimidating complexity of legal procedures. Traditional approaches to legal consultation typically demand significant time commitments, substantial financial resources, and physical presence at courts or law offices—barriers that disproportionately affect vulnerable populations. This paper introduces NyaySetu, a thoughtfully designed system that combines artificial intelligence with practical legal guidance. The system allows users to describe their legal concerns in everyday language, then uses natural language processing techniques to understand the query and machine learning algorithms to identify relevant legal categories. Beyond simple classification, the system offers practical suggestions and helps users connect with appropriate legal professionals for further assistance. By making preliminary legal guidance more accessible while reducing unnecessary consultations, the system addresses genuine needs in legal service delivery. The approach maintains careful ethical boundaries by offering informational support rather than legal advice, ensuring users understand when professional assistance remains necessary. Experimental evaluation confirms that the classification models perform effectively, suggesting the system can scale to serve diverse user populations across different legal contexts.

KEYWORDS - Artificial Intelligence, Natural Language Processing, Legal Tech, Chatbot, Machine Learning, Case Classification, Legal Assistance

1. INTRODUCTION

The ability to access legal help when needed forms a cornerstone of fair and just societies. Yet for countless individuals, obtaining basic legal guidance presents overwhelming obstacles. Legal fees often exceed what ordinary people can afford, legal language remains inaccessible to those without specialized training, and even understanding what type of lawyer to consult can become a confusing first hurdle. These difficulties lead many to delay addressing legal matters or to navigate complex situations without any guidance, often with negative consequences.

Technology has transformed how people access information across nearly every domain of life, yet legal services have remained surprisingly resistant to such change. While countless applications help people manage finances, track health, or navigate unfamiliar cities, comparable tools for understanding legal situations have remained limited. The few digital legal resources that exist often assume users already understand their legal situation well enough to know what questions to ask or what services to seek.

The gap between available legal technology and genuine user needs becomes particularly apparent when considering how people actually approach legal problems. Most individuals cannot articulate their situation using precise legal terminology, nor should they need to. They simply know something has gone wrong—a landlord refuses to return a deposit, an employer has withheld wages, a family dispute

has escalated—and they need someone to help them understand what options exist.

NyaySetu was developed with this reality firmly in mind. The system aims to meet users where they are, allowing them to describe their situation in ordinary terms, then working to make sense of what they share. By combining the pattern-recognition capabilities of machine learning with structured pathways to professional consultation, the system attempts to create a bridge between everyday legal concerns and meaningful assistance.

The work described here pursues several connected goals:

- Building a system that can interpret legal questions expressed in natural, non-technical language
- Developing reliable methods for categorizing legal issues based on user descriptions
- Creating practical pathways that help users understand their situation without unnecessary expense
- Ensuring the system remains accessible to individuals regardless of their prior legal knowledge
- Maintaining ethical standards by clearly distinguishing informational support from legal advice

2. RELATED WORK

The intersection of computing and legal practice has drawn increasing research attention over the past decade, with

investigators exploring how artificial intelligence might support various aspects of legal work. This body of work provides important foundations while also revealing persistent gaps that motivated the current project.

2.1 Text Classification in Legal Contexts

Researchers have made considerable progress applying machine learning methods to legal text classification. Early work demonstrated that relatively straightforward algorithms, including Naive Bayes classifiers and support vector machines, could distinguish between broad legal categories such as criminal, civil, and family law based on textual features extracted from case documents. These successes suggested that automated categorization could support various legal applications.

More recent investigations have explored deep learning approaches, with neural network architectures showing improved ability to capture contextual relationships within legal texts. Transformer-based models, particularly BERT and variants fine-tuned on legal corpora, have demonstrated notable advances in understanding the nuanced language characteristic of legal writing. These models can identify relevant legal concepts across extended passages and recognize subtle distinctions between similar case types.

2.2 Conversational Systems for Legal Assistance

The development of conversational agents for legal contexts has followed multiple design approaches. Rule-based systems, which follow explicitly programmed decision trees, offer the advantage of predictable, controllable responses within well-defined domains. However, they struggle when users ask questions that fall outside anticipated patterns.

Retrieval-based systems match user queries against databases of previously answered questions, providing reliable responses for common inquiries but failing when encountering novel or complex situations. Generative approaches, which produce original responses using language models, offer greater flexibility but require careful oversight to prevent the generation of misleading or inaccurate information.

2.3 Digital Platforms for Legal Consultation

Commercial and nonprofit organizations have launched various platforms intended to connect individuals with legal professionals. These platforms typically offer lawyer directories, scheduling tools, and video consultation features. While they serve important functions, they generally assume users already understand enough about their legal situation to select appropriate representation. This assumption creates a significant barrier for precisely the users who might benefit most from such services.

2.4 Identified Gaps

Analysis of existing work reveals several limitations that current systems have not adequately addressed. Most available tools focus either on document analysis or on consultation matching, with few attempting to bridge these functions in integrated ways. Personalization remains limited,

with systems often providing generic guidance regardless of user circumstances. Real-time interaction capabilities, while common in other domains, remain underdeveloped in legal contexts. Perhaps most significantly, existing systems rarely address the economic constraints that make legal services inaccessible for many potential users, particularly students and those with limited incomes.

These gaps suggest the need for systems designed differently—systems that combine understanding of user language with meaningful pathways to assistance, that remain accessible without sacrificing accuracy, and that acknowledge the genuine needs of users approaching legal questions from positions of limited prior knowledge.

3. PROBLEM STATEMENT

The challenges that prevent people from accessing legal assistance operate at multiple levels, creating compounded barriers that many individuals cannot overcome without help.

Knowledge Barriers: People facing legal situations often cannot determine whether their situation actually involves legal questions, what laws might apply, what options exist, or what type of professional could help. This uncertainty frequently leads to inaction or misguided efforts that waste time and resources.

Economic Barriers: Legal consultation fees present immediate obstacles for individuals with limited financial resources. Hourly rates that seem reasonable to legal professionals appear prohibitive to those living paycheck to paycheck, leading many to forgo consultation altogether even when their situations warrant professional attention.

Geographic Barriers: Physical access to legal services remains unevenly distributed, with rural areas and economically disadvantaged communities often having few local practitioners. Travel requirements add both expense and logistical complexity to what might otherwise be straightforward consultations.

Procedural Barriers: Legal systems operate according to rules that appear arbitrary to outsiders. Filing requirements, documentation standards, and court procedures create pitfalls that can derail legitimate claims before they receive substantive consideration.

Information Fragmentation: Available resources that might help—legal aid organizations, self-help materials, bar association referral services—operate independently, requiring users to discover, evaluate, and navigate multiple systems without cohesive guidance.

These interconnected challenges suggest that piecemeal solutions will prove insufficient. What appears needed instead is a system that addresses the entire journey from initial confusion to appropriate assistance, providing guidance at each step while recognizing when professional help becomes necessary.

4. RESEARCH METHODOLOGY

The development of NyaySetu followed a structured process designed to create a system that genuinely serves user needs while maintaining technical reliability and ethical standards.

4.1 Overall System Design

NyaySetu employs a three-component architecture that processes user input through sequential stages:

The conversational interface component accepts user descriptions in natural language, performs basic linguistic processing to identify key terms and concepts, and prepares structured representations suitable for analysis. This component prioritizes flexibility, allowing users to describe their situations in whatever terms feel natural rather than forcing them into rigid form structures.

The classification component uses machine learning models trained on legal datasets to identify likely legal categories and case types based on the processed user input. Rather than making definitive judgments, the system generates probability distributions across categories, reflecting the inherent uncertainty of preliminary analysis.

The guidance and connection component translates classification results into practical suggestions presented in accessible language. When appropriate, it recommends consultation with legal professionals who specialize in relevant areas and provides mechanisms for initiating such consultations directly through the platform.

4.2 Data Collection and Preparation

Training effective classification models required assembling datasets representative of the legal questions users might present. Data was gathered from multiple sources:

- Public legal databases provided case descriptions with established category labels, offering reliable examples of how legal issues are formally described. Legal aid organizations contributed anonymized records of actual client inquiries, providing valuable examples of how ordinary people describe legal problems in their own words. Simulated user queries were developed to cover categories underrepresented in existing datasets, ensuring the training data captured the full range of potential user inputs.
- All collected text underwent preprocessing to improve consistency while preserving meaningful content. This processing included tokenization to separate text into individual units, removal of terms unlikely to contribute to accurate classification, normalization to address variations in word forms, and filtering to eliminate irrelevant characters or formatting artifacts.

4.3 Feature Engineering and Model Selection

Converting text into forms suitable for machine learning required careful feature selection. Term frequency-inverse document frequency (TF-IDF) weighting was employed to capture which terms appear most distinctive across different legal categories. This approach helps the model focus on language that meaningfully distinguishes between case types. Multiple classification algorithms were evaluated to identify approaches that balance accuracy with practical deployment considerations. Naive Bayes classifiers offered computational efficiency and straightforward implementation. Support vector machines provided stronger

performance on certain classification tasks but required more careful parameter tuning. Ensemble approaches combining multiple algorithms were explored to leverage strengths across different classification scenarios.

4.4 Model Training and Evaluation

Models were trained using supervised learning approaches, with labelled datasets providing examples of correct classifications. Training data was divided into development and test sets to evaluate performance on previously unseen examples. Cross-validation techniques provided additional robustness by testing models across multiple data partitions. Evaluation metrics focused on classification accuracy across legal domains and specific case types. Particular attention was paid to performance on the kinds of informally phrased queries that characterize actual user input, recognizing that models often perform differently on carefully written legal texts versus everyday language.

4.5 Ethical Considerations

Several ethical principles guided system development. The system explicitly avoids providing direct legal advice, instead offering informational guidance and suggesting appropriate professional consultation. Clear disclaimers inform users about system limitations and emphasize when professional assistance remains necessary.

User privacy received careful attention throughout design, with the system collecting only information necessary for classification and consultation matching. Data handling practices prioritize user control and minimize retention of sensitive information.

5. RESEARCH GAPS

While existing legal technology has made meaningful contributions, careful examination reveals persistent gaps that current approaches have not adequately addressed.

Integration Deficit: Most available tools focus on either automated analysis or consultation matching, rarely combining these functions in ways that serve users throughout their journey from initial confusion to appropriate assistance. This fragmentation forces users to navigate multiple systems without coordinated guidance.

Personalization Limitations: Existing systems typically provide the same responses regardless of user context, failing to account for factors such as jurisdiction, user resources, or specific circumstances that might affect appropriate courses of action.

Real-Time Capability Gaps: Despite widespread expectations for immediate digital interaction, many legal resources require users to wait hours or days for responses, limiting their utility when users need timely guidance.

Economic Accessibility Oversights: Systems rarely consider the financial constraints that make legal services inaccessible for many users, failing to incorporate features that might help users understand costs or identify lower-cost alternatives.

Ethical Framework Inconsistencies: Approaches to managing the boundaries between information and advice vary widely, with some systems potentially overstepping

appropriate limits while others provide insufficient guidance to be genuinely useful.

These gaps suggest opportunities for systems designed differently—systems that integrate rather than fragment, personalize rather than generalize, respond rather than delay, and acknowledge rather than ignore the economic realities of their users.

6. DISCUSSION

The development process revealed several important considerations that shaped final system design and offer lessons for similar projects.

Data Quality Challenges: Assembling adequate training data proved more challenging than initially anticipated. Public legal datasets, while valuable, often employ language quite different from how ordinary users describe their situations. Legal aid records provided more realistic examples but required substantial cleaning to remove identifying information and standardize formats. The project ultimately employed a hybrid approach, combining multiple data sources with targeted data generation to ensure coverage across legal categories.

Model Performance Considerations: Classification models demonstrated strong performance on well-structured inputs but showed more variable results on informally phrased queries. This finding reinforces the importance of careful interface design that helps users provide sufficient information without requiring legal expertise. The system incorporates clarifying prompts that guide users toward providing complete descriptions without imposing rigid structures.

Balancing Accuracy and Accessibility: Trade-offs between classification precision and user experience required careful navigation. Overly narrow models might produce accurate classifications but fail when user inputs deviate from training examples. Broader models offer greater flexibility but risk less precise categorization. The chosen approach employs confidence thresholds that trigger clarifying questions when classifications lack certainty, preserving accuracy without sacrificing accessibility.

Ethical Implementation Complexities: Maintaining appropriate boundaries between information and advice required continuous attention throughout development. Language that seems clearly informational to developers might be interpreted as advice by users seeking guidance. The system incorporates multiple mechanisms to manage this challenge, including explicit disclaimers, careful phrasing of responses, and clear pathways to professional consultation when situations warrant.

User Trust and System Adoption: Early feedback suggests that user trust depends on multiple factors beyond technical accuracy. Transparency about system limitations, consistency in responses, and seamless transitions to human consultation all contribute to user confidence. The system design prioritizes these factors alongside classification performance.

7. PROPOSED SYSTEM ARCHITECTURE

NyaySetu implements a modular architecture designed for flexibility, scalability, and maintainability.

User Interface Layer: The conversational interface presents a clean, accessible design that guides users through describing their situations without requiring legal knowledge. Input can be provided through typing or, in future implementations, through voice recognition. The interface adapts to user responses, asking clarifying questions when needed and providing progressive guidance as situations become clearer.

Processing Layer: User input flows to processing components that perform linguistic analysis, extract relevant concepts, and structure information for classification. This layer employs natural language processing techniques including tokenization, part-of-speech tagging, and entity recognition tailored to legal contexts.

Classification Layer: Structured representations pass to machine learning models that generate probability distributions across legal categories and case types. Multiple models operate in parallel, with ensemble techniques combining their outputs to improve reliability.

Guidance Generation Layer: Classification outputs inform the generation of user-appropriate responses. The system provides plain-language explanations of likely legal categories, outlines common options for proceeding, and identifies situations where professional consultation is strongly recommended.

Consultation Integration Layer: When users indicate interest in professional consultation, the system accesses a directory of legal practitioners filtered by relevant specializations and geographic availability. Scheduling tools and video consultation features enable immediate connection without requiring separate platform navigation.

Data Management Layer: User interactions, classification histories, and consultation records are managed with attention to privacy and security requirements. Data retention policies prioritize user control and minimize storage of sensitive information.

8. FUTURE DIRECTIONS

Several promising directions for extending and improving the system merit consideration.

Advanced Language Models: Incorporating more sophisticated language models, particularly legal-domain adaptations of transformer architectures, could improve classification accuracy and response quality. These models offer enhanced ability to capture subtle legal distinctions and contextual relationships within user descriptions.

Multilingual Capabilities: Extending the system to support multiple languages would dramatically expand accessibility, particularly in regions where English proficiency cannot be assumed. Initial focus might include languages with significant speaker populations currently underserved by legal technology.

Voice-Based Interaction: Voice interfaces could reduce barriers for users who struggle with typing or prefer conversational interaction. This approach might also support accessibility for users with visual impairments or literacy challenges.

Legal Resource Integration: Direct connections to relevant legal documents, forms, and procedural guides could provide users with practical resources matched to their identified

situations. Such integration would extend the system beyond guidance into actionable support.

Mobile Application Development: Native mobile applications could improve accessibility by enabling interaction from any location, supporting offline functionality where connectivity is limited, and leveraging device capabilities for enhanced user experience.

Continuous Learning Mechanisms: Implementing feedback loops that improve system performance based on user interactions would enable ongoing refinement. Mechanisms for identifying classification errors, collecting user corrections, and incorporating new legal developments could support sustained improvement.

9. CONCLUSION

The work presented here demonstrates the feasibility of creating integrated systems that combine AI-powered legal guidance with practical pathways to professional consultation. NyaySetu addresses genuine challenges in legal service accessibility by meeting users where they are—allowing descriptions in ordinary language, providing meaningful preliminary guidance, and facilitating connections to appropriate practitioners when needed.

Several principles guided development throughout. Accessibility required systems that work for users regardless of prior legal knowledge. Practical utility demanded outputs that translate into actual assistance rather than abstract information. Ethical responsibility necessitated clear boundaries between information and advice, with explicit guidance about when professional consultation becomes necessary.

The resulting system represents a step toward legal technology that serves rather than confuses, that guides rather than directs, and that connects rather than fragments. While much work remains—in improving classification capabilities, expanding language support, and refining user experience—the foundation established here suggests promising pathways forward.

More broadly, this work illustrates how thoughtful application of artificial intelligence can address genuine social needs when technical development proceeds in close connection with user realities. Legal access challenges will not disappear overnight, but systems that help individuals understand their situations, explore their options, and connect with appropriate assistance can meaningfully contribute to more equitable outcomes.

The journey from initial user confusion to appropriate assistance remains complex, with many factors beyond any single system's control. Yet by carefully designing each component to serve real needs, by maintaining clear ethical boundaries, and by remaining responsive to how users actually interact, systems like NyaySetu can become valuable tools in the larger effort to make legal assistance accessible to all who need it.

10. REFERENCES

[1] [1] D. Jurafsky and J. H. Martin, *Speech and Language Processing*, 3rd ed. Pearson, 2023.

- [2] [2] F. Sebastiani, "Machine learning in automated text categorization," *ACM Computing Surveys*, vol. 34, no. 1, pp. 1–47, 2002.
- [3] [3] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*. O'Reilly Media, 2009.
- [4] [4] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of NAACL-HLT*, 2019, pp. 4171–4186.
- [5] [5] I. Chalkidis, M. Fergadiotis, P. Malakasiotis, N. Aletras, and I. Androutsopoulos, "Legal-BERT: The muppets straight out of law school," in *Proceedings of EMNLP*, 2020, pp. 2898–2904.
- [6] [6] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [7] [7] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *Proceedings of ICLR Workshop*, 2013.
- [8] [8] A. Vaswani et al., "Attention is all you need," in *Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008.
- [9] [9] K. S. K. S. N. Kumar and A. Singh, "Detecting smishing attacks using NLP and machine learning," *IEEE Transactions on Cybersecurity*, vol. 8, no. 3, pp. 245–258, 2024.
- [10] [10] P. Verma and S. Rao, "Privacy assessment in Android apps: A systematic review," *Computers & Security*, vol. 108, 2021.
- [11] [11] R. Ahmed and X. Li, "Subscription management and user engagement analytics," *Journal of Digital Services*, vol. 14, no. 2, pp. 112–128, 2020.
- [12] [12] V. Choudhary and M. Patel, "Automated privacy requirement analysis in mobile apps," *IEEE Software*, vol. 34, no. 4, pp. 48–55, 2017.
- [13] [13] M. Lewis and R. Perry, "Privacy challenges in mobile finance apps," *Journal of Mobile Systems*, vol. 11, no. 3, pp. 178–192, 2019.
- [14] [14] P. Bitrián, I. Buil, and S. Catalán, "Gamification and engagement in financial apps," *Information Systems Research*, vol. 32, no. 4, pp. 1123–1141, 2021.
- [15] [15] A. Alenazi and C. Sas, "Limitations of current financial tracking systems," *Journal of Finance and Technology*, vol. 7, no. 2, pp. 89–104, 2023.