iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** XI **Month of publication:** November 2025

DOI: https://doi.org/10.22214/ijraset.2025.74954

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

NyayaSahaya: A Survey on AI-Powered Legal Assistance and Government Policy Simplification for Common People

Abhishek Pawaskar[1], Tejas Kunde[2], Yash Sawant[3], Prof. Deepali Shrikhande[4]
[1, 2, 3, 4]*Department of Information and Technology, Vidyalankar Institute of Technology Mumbai, Maharashtra, India*
[4]*Department of and Technology, Vidyalankar Institute of Technology Mumbai, Maharashtra, India*

Abstract: In developing nations, legal literacy continues to be a major challenge because the jargon employed in legal discourse and state policies serves as barriers to everyday citizens seeking legal help. This survey considers the space of AI-based legal support systems, particularly the document simplification and multilingual natural language processing approaches. We outline the existing legal AI systems deploying chatbots, contract analysis software, and policy simplification systems, their technical approaches, and limitations. The major challenges found include: data limitations within the legal space, the challenges of processing multilingual content, and accuracy of legal advice generated by AI. We introduce NyayaSahaya as a combined legal advice generator and government policy simplifier utility, aimed at non-technical users operating in multilingual environments. This survey adds to the understanding of AI systems which can through value and access democratize legal aid in developing nations who would likely require digital governance development efforts. Recent reports of legal professional uptake of AI systems indicate that in the face of legal literacy challenges, there is impetus for widespread adoption of these technologies by 2024 (increase to up to seventy-nine percent of legal professionals)
Keywords: Legal AI, Natural Language Processing, Text Simplification, Multilingual NLP, Digital Governance, Legal Chatbots.

I. INTRODUCTION

A. Problem Statement and Motivation

Legal literacy is a huge problem in developing countries where around seventy percent of the population does not possess even a rudimentary understanding of their own legal rights and the government's stated policies[1] The multiplicity of legal terminology and the economic reasons for not being able to afford legal representation creates a chasm between legal systems and the people they are intended to serve. In nations like India, legal policies are often printed in obscure bureaucratic vernacular which the rural and economically disadvantaged class of the nation cannot interpret. Even policies pertaining to agricultural subsidy schemes, healthcare benefits, or housing grants are increasingly under-utilized, as the average person does not understand the meaning of a vernacular filled with professionalism and legalese, resulting in lost opportunities for millions.

The standard legal service model presents a number of restrictions that together cause legal exclusion. The cost of consultations typically ranges from $100 to $300 per hour, which is not affordable for most citizens. Language barriers also impede access to legal documents since the vast majority of legal documents are only available in English even in highly diverse populations. Geographically, access to experienced legal professionals is limited and sometimes impossible even in a rural area with little legal infrastructure. These lengthy processes prevent citizens from seeking help and leave them unaware of government assistance, denying them the ability to safeguard their rights, or forced to make uninformed legal decisions that may never be reversed.

These issues have been exacerbated by the post-pandemic era, with surging demand for remote, inexpensive, and immediate legal services most evident in underserved populations [4]. Recent technological advances in artificial intelligence and natural language processing offer possible solutions to bridge the accessibility gap [16, 17]. An AI-powered legal assistance system can provide continuous availability, low-cost consultation, multilingual capability, and simple explanations of complex legal concepts. However, the development of such systems requires careful attention to multiple factors related to the accuracy of legal advice, cultural responsiveness, and ethical considerations. It is critical that AI systems do not misunderstand local laws or provide misleading advice, as this can have disastrous consequences for users who rely on that advice. As AI tools are developed, they need to accommodate a variety of end-user issues, and make certain that individuals with limited education or technology experience can access services without it resulting in misinformation or worse, legal harm.

B. Research Questions and Scope

This study aims to examine several interconnected research questions that shape the analysis of AI-augmented legal advice systems. First, we investigate existing methods for legal text processing and advice generation using artificial intelligence and machine learning, assessing how current architectures and models perform in practical legal contexts. Second, we evaluate how effectively today's legal AI systems serve non-expert users—considering usability, accuracy, and accessibility. Third, we analyze the challenges involved in simplifying complex legal and policy documents through natural language processing, with the goal of making legal information more understandable to the general public. Finally, we explore how socio-economic and cultural factors influence the accessibility and adoption of AI-driven legal assistance tools.

The scope of this research focuses on English-language AI-powered legal support systems. Our emphasis is on tools designed to assist everyday users rather than legal professionals, prioritizing accessibility, usability, and fairness in information delivery. While multilingual systems represent an important future direction, this study concentrates exclusively on English-based implementations to maintain consistency and technical feasibility. We also examine how such AI systems can align with digital governance initiatives, enhancing citizen engagement and compliance by making legal information more approachable and transparent.

II. LITERATURE REVIEW

A. AI in Legal Domain

The utilization of artificial intelligence within legal services has progressed tremendously over the last ten years - from rudimentary document research systems to advanced conversational agents capable of offering actionable legal advice. The launch of DoNotPay, in 2015, represents one of the first legal chatbots to succeed and originally began as a bot for disputing parking tickets, expanding to consumer legal issues from small claims disputes to lease agreement review [2] . Processing over one hundred sixty thousand legal requests per month [2] DoNotPay exemplifies the demand for a new way of providing legal services that exists outside the traditional legal services channel.

ROSS Intelligence was the first to harness IBM Watson's natural language processing capabilities to carry out AI-driven legal research. Its goal was to enhance the effectiveness of legal research and to reduce the amount of time and effort lawyers spent on the task . ROSS Intelligence claimed to achieve about ninety percent accuracy—the type of accuracy lawyers had come to expect from traditional keyword-based legal research, although the company acknowledged some difficulty in complex legal reasoning and context-specific legal analysis that took into account much deeper semantic analysis. Other services, such as Thomson Reuters' Westlaw Precision, have also leveraged generative AI to improve search capabilities with a specific focus on leveraging the search capabilities of iterative, generative AI, as Westlaw Precision allows lawyers to ask thought-provoking questions of the AI and receive instant legal answers. In some instances, Westlaw purported that this dramatically improves lawyers' efficiencies in legal research by approximately forty percent .

In recent years, advancements like LawBot and other conversational legal systems have relied on transformer-based architectures, which are a step in the direction of improved natural language understanding . These systems seem to be able to classify intents with seventy-five percent to eighty-five percent accuracy when posed with common legal questions, although their accuracy will vary based upon specific legal contexts or jurisdictions. Furthermore, Clio's 2024 Legal Trends Report confirms that a recent increase in AI adoption in legal practice has spiked to seventy-nine percent overall, with evaluative tools supporting the review of documents, client communication, and case organization becoming more established .

Contract analysis is another important application area where AI shows a lot of promise. Companies like Kira Systems and LawGeex have achieved "human-level" performance on contract review tasks, with one study finding that LawGeex was ninety-four percent accurate in identifying legal issues in non-disclosure agreements, compared to eighty-five percent for human attorneys with several years of experience [5]. Recently, tools like Harvey and Paxton AI have taken this further, incorporating multimodal analysis that combines the text in contracts with visuals such as tables or signatures to improve accuracy and usability . These developments suggest the opportunity for AI to simplify legal analysis workflows, despite questions about liability and professional supervision.

B. Document Simplification and Summarization

Simplifying text in legal contexts involves challenges due to the precision of legal language and the severe consequences of oversimplification. Earlier approaches concentrated on lexical simplification, and taught to replace complex legal terms with simpler alternatives [6]. For example, the term "indemnification" can be simplified to "protection," but may overwrite an important legal meaning with respect to liability or obligation, which can significantly affect legal interpretation.

This challenge becomes especially pronounced in multilingual contexts, where the intended legal meaning of a word may miss the mark during simple translation, due to differences between cultural and legal frameworks.

Over the years, neural approaches to text simplification have exhibited even more favorable outcomes. The use of sequence-to-sequence models trained on parallelized corpora of complex texts and simplified legal texts achieved BLEU scores ranging from 0.65 to 0.72 on legal document simplification exercises . Accuracy of coherent simplified outputs that respect semantic meaning have been shown to have better outcomes with transformer-based models, especially when utilizing the fine-tuned variants of T5 and BART trained on legal texts. Other research has considered hybrid approaches and human-in-the-loop models which provide iterative reformulation of simplifications by using user feedback loops to correspondings with user understanding levels, rather than one-size-fits-all transformation process [28].

Legal text summarization is particularly challenging due to its hierarchical structures and inter-text references (for example legal documents may cite or reference other legal documents in full or partially). Extractive summarization with attention-based models has achieved ROUGE-L scores of 0.58 to 0.62 for legal case summarization [8]. Abstractive summarization is more difficult to implement consistently, but is easier for non-expert users to read since it creates concise summaries that resemble human summaries with essential information and little detail. The AI summarizer created by LexisNexis can perform multi-document summaries by extracting information across multiple case law, statutes, and regulatory law documents .

Recent research on the simplification of government policy has explored the use of multi-document summarisation techniques that can process full policy frameworks and generate explanations suitable for citizens [9]. Typically, these systems employ hierarchical attention models to surface relevant policy points and benefits to citizen sub-populations (e.g. simplifying tax policies or welfare schemes into bullet point summaries). Simplifying government information has been shown to improve engagement with government programs, and pilot studies show that if government use a simplified policy transcript this can improve participation in programs by 25% when such materials are simplified . In summary, an AI-based simplification model demonstrates its value between complicated government language and citizen understanding when summarising complex documents.

C. Multilingual NLP for Legal Applications

Legal settings have language and jurisdictional variances that present unique challenges for multilingual processing. Much of the focus in cross-lingual legal information processing has been on European languages, with very little attention to processing of Indian languages until recently [10]. Recent strides toward NLP in India have begun to lessen this gap with multilingual models aimed specifically at Hindi, Tamil, Bengali, and other regional languages to assist in tasks like legal query processing and contract analysis [11]. These are important steps in providing legal access in multilingual communities, where language barriers often prevent residents from understanding their rights or services available to them.

Methods using transfer learning have shown promise in thinking about low-resource legal language processing. Multilingual BERT variants help with 70% - 80% accuracy on legal entity recognition tasks in Indian languages like Hindi and Marathi, but accuracy falls far below English baselines as a result of a lack of annotated training data and the intricate legal vocabulary used in regional languages. Building new datasets for legal NLP tasks such as the Indian Legal NLP Benchmark contain the potential to close that gap by providing some much-needed resources for tasks like judgment prediction, entity extraction, and legal question answering, which, in turn, allow for more resilient model development .

In an Indian legal context, code-switching presents distinctive challenges, as conversations often contain English legal terms that are intermixed with explanations in their local language utilizing their regular speech patterns and combinations of languages. Recent research into code-switched legal queries achieves sixty-five percent accuracy in intent classification , which reflects the technical obstacles for systems to fully respond to complex linguistic features . Recent advances in this area have been to develop hybrid models that perceive code-mixing and improve performance interaction when the user toggles back and forth between languages and formal and informal legal speech. For example, systems that accept legal queries such as "What is the bail process in Hindi and English?" have gained increased following among users who have traditionally not experienced formal legal education while having formal legal needs.

Cross-lingual embeddings have been used for the matching of legal documents across languages allowing citizens to access relevant legal information irrespective of the language of origin. Government policy analysis systems are increasingly portalizing multi-linguicism, however most remain limited to major Indian languages. Initiatives such as IndiaAI's initiatives are broadening this range to include less-resourced languages like Assamese and Odia which is essential for equitable access . All of this is important to make AI legal tools accessible to varied populations more equally, especially in federal systems as language diversity often correlates adversitely with socioeconomic access.

III. TECHNICAL APPROACHES

A. Core NLP Techniques

Legal entity recognition is at the core of most legal AI systems because it involves identifying parties, dates, legal citations, and jurisdictions in unstructured text. Named Entity Recognition models trained on legal corpora achieve F1 scores between 0.85 and 0.92 on legal entity types . While earlier models were primarily based on conditional random fields, recent models like Legal BERT employ transformers and are pre-trained on a legal corpus. These models are particularly useful for extracting the entity types described above in court judgments or contracts and are essential for downstream legal reasoning tasks that involve identifying case numbers, statutory citations, and party designations. Extracting information from legal documents involves advanced parsing methods that deal with the complex structures, nested clauses, and tables of contents of legal documents. Recent work has used hierarchical attention mechanisms and graph neural networks to incorporate document structure and the relationships between legal concepts [14] into modern methods [14]. This also allows the system to produce a clear record of the implied dependencies, explicitly or implied by earlier text, e.g., which clauses or cases influence others within a legal agreement. Take, for instance, the task of extracting conditions from a tenancy agreement. In addition to terms expressly contained in the tenancy contract, sub-contracting clauses may impose obligations that can impact others, e.g. by way of conventions for legal interpretation that are inherent within relationships of the parties involved. Legal question answering systems encounter distinctive challenges of reasoning over legal knowledge, while being accurate and managing hallucination. The most common approaches involve retrieval-augmented generation approaches that typically extract passages through dense passage retrieval before using those passages to generate an answer utilizing a transformer. These systems currently achieve 78% exact match accuracy on legal QA benchmarks, although accuracy is strongly variable across legal domains [15]. Future improvements for these systems include utilizing external knowledge bases (like government policy databases) as base knowledge with the hope of improin accuracy in novel scenarios and improve contetual relevance

B. ML and DL Architectures

In the last few years, transformer infrastructures have completely reshaped how to approach legal text processing with models built for the legal domain, such as LegalBERT and CaseLaw-BERT, outperforming the general-purpose model. LegalBERT, for example, was pre-trained with twelve gigabytes of legal text, demonstrating a fifteen- to twenty-percent performance boost as compared to BERT-base on legal case classification . More recent versions leverage even large sized pre-training corpora, many of which involve international case law, regulatory texts, and other institutional resources strategically used to help the models better generalize across jurisdictions. Professional models do far better at understanding legal vocabulary, reasoning styles, and document organization compared to general language models that pre-trained on vast corpora that were not exclusively, or prioritizing, legal corpus. Large language models offer new possibilities and new challenges for use in law. GPT-3 and other large models have demonstrated strong performance on tasks requiring legal reasoning, and they have significant issues with hallucination and inconsistency, which can be troublesome in law and legal practice where accuracy is very important [17]. Fine tuning this type of models with legal data sets is a promising way to try to minimize reliability issues, although this model fine-tuning introduces new challenges of proper validation to ensure outputs comply with professional standards. More recently, GPT-4 has been modified for legal use while avoiding making up sample rules of law or cases, and has been shown to be at least eighty-five percent accurate on sophisticated legal reasoning tasks [33]. Concerns about liability and professional responsibility still exist, and pose a challenge due to uncertainty about legal models proper to assure accuracy and the competence to practice law. Sequence-to-sequence models for the generation of legal documents use encoder-decoder architectures to process and signficantly improve fluency of a legal document using attention to features of long-range dependency present in legal text. T5-based models that are fine-tuned with pairs of legal documents have shown comparably state-of-the art performance in legal document simplification, producing BLEU scores that exceed 0.70 as compared to the existing benchmarks . Further, these simplified legal documents models are commonly extended with reinforcement learning in order to first prioritize accuracy over fluency, while a second objective attempts to improve readability in later sequential returns. This work tries to strike a balance between two aspects of legal document processing, accessibility and precision. Cross-lingual methods for multilingual legal processing generally use multilingual transformer models with language-specific fine-tuning. Recent work has shown that language-specific legal pre-training improves performance on downstream tasks, even when the legal specific training/supervision data is small . For Indian contexts, the approach is informed by pre-training on a mixed language corpus, in order to take into account real-world use such as code-switching in legal consultations, etc. In addition, new models have been developed that include dialect-specific embedding to account for linguistic variations in regional languages that impact meaning and interpretation, which is critical in legal contexts where precision is required

IV. KEY CHALLENGES AND LIMITATIONS

A. Technical Challenges

The biggest barrier to developing legal AI systems is data scarcity, particularly when it comes to specialized domains and low-resource languages. In the legal sector, datasets can be proprietary, sensitive or simply limited to allow for training robust models that generalize well across legal contexts [21]. Public legal datasets generally focus on case law rather than practical legal advice scenarios that may be relevant to common citizens, which can result in overfitting - poor performance for eventualities - and can lead to a situation where models simply don't learn or generalize to core concerns. For instance, datasets in the consumer dispute realm are far less robust than datasets in the corporate litigation realm, which means the very populations in need of accessible legal assistance are at a detriment and imbalance.

Understanding complex legal language represents a significant impediment for natural language processing systems. Legal documents utilize specialized vocabulary, often complicated sentence structure, and implicit references requiring deep domain consensus to deconstruct. For example, resolving ambiguity in a legal context can stem from common legal knowledge, often based in precedent or jurisdictional differences - matters which are nearly impossible to convert into an AI system without extensive legal experience. Additionally, hallucinations created by generative models compound the challenge; one study found that AI legal tools fabricate facts in up to seventeen percent of queries raising serious concerns for users relying on the information to make legal decisions .

The issues of multilingual processing for the Indian languages is compounded beyond just translation. Technical limitations created by limited resources, location of scripts, and multi- lingual diversity limit standard multilingual models' capabilities. Code-switching between English and local languages in legal dialogues presents challenges for parsing, while transliteration differences create additional obstacles for entity matching and recognition [21] Secondly, there could be variances in legal meaning and interpretations depends on the province adding layers of complexity and necessitating a systems flexibility integrating local legal systems. For example, there will be substantive differences in property laws in Maharashtra and Kerala which will necessitate context-sensitive processing capabilities to consider the rules associated with a particular jurisdiction.

Reliability and consistency of models continue to be an important issues when it comes to providing legal AI systems with reliableconsistency in similarity situations. These legal AI applications generally give different answers to the same queries than similar queries, which can significantly undermine consumer confidence in those responses and potentially contribute to be disastrous decisions. Recent cybersecurity challenges with legal AI platforms, such as leaks of each company's respective confidential information also create additional concerns requiring robust encryption and anonymization protocols [35], and need to involve more encryption and anonymization of legal data. All of these technical considerations need to be addressed before widespread adoption of legal AI with trust from the public.

B. Domain and Ethical Challenges

The prevailing legal requirement for accuracy surrounding AI-generated advice presents major liability issues. While there may be a tolerable or correctable level of error with other applications of AI, legal advice applications must be highly accurate to avoid causing the user harm through inaccurate advice. As current systems only demonstrate accuracy levels between seventy and eighty-five percent on legal advice generation tasks, this level of accuracy may not be acceptable for life decisions that may have negative financial consequences, lead to imprisonment or loss of rights . The determination of liability for AI incorrectly providing legal advice is still an unanswered question posing a barrier for current and future AI legal advice applications as the legal ramifications of unclear liabilities are not clarified between developers, platforms, and users. Jurisdictional differences reveal complex challenges that multiply through federal and state jurisdictions. Legal systems differ from state to state, country to country, and local jurisdiction to local jurisdiction so that AI will need to address the locality's variations in laws, procedures, and cultural practices through localizing and adapting AI. In a system of federalism, like India's, fields of law such as property, marriage, or local taxation might vary significantly at the state level [21]. An AI that provides legal advice about property rights must understand the national statutes, state statutes, and local practices that may influence the practical or legal interpretation.

The question of bias in legally oriented AI systems raises significant ethical questions that will more significantly impact marginalized communities likely to use AI-assisted legal services as a result of their socio-economic standing. Research has shown that advanced systems may demonstrate systematic bias in how they approach specific demographics and fields of law [23]. To illustrate the point, consider AI being trained on a dataset that is biased toward cases occurring in urban environments; this could potentially disadvantage rural users and reinforce existing inequities in access to, and outcomes of, legal services. In addition to the potential bias, this form of automated service can lead to unjust treatment when disadvantaged individuals receive inequitable and ineffective legal discrepancy due to the automation, potentially aggravating access to justice and overall representation in the future.

Concerns around privacy arise from the provision of personal information regarding legal matters in trust with AI systems. Users may be less willing to provide access to their legal data, particularly with outside individuals or systems, due to concerns related to data security and similar apprehensions regarding potential third-party or government uses of what was provided. For example, a regulatory burden established by the Digital Personal Data Protection Act in India has created compliance requirements for robust data protection measures [36] to institute additional protective measures, including encryption, anonymization, and mechanisms to obtain consent from users allowing the data to be retained by the AI systems [36] Finding the balance between the need to collect the data to improve the overall performance of the systems and the individual user's privacy rights remains an ethical concern for those developing legal AI applications.

Cultural and socioeconomic differences affect the way in which legal advice is interpreted and applied in different social groups. AI systems must account for the local customs, economic realities, and social factors influencing legal decision making within different communities, so it goes beyond formal law to consider what is possible. For example, recommending litigation may be impractical for users who cannot afford court fees or do not have access to legal representation - context appropriate solutions must take the reality of the users situation into account . Accordingly, effective legal AI systems cannot ignore an awareness of socioeconomic factors, so that the recommendations can be legally valid, and possible to address the user's lived experience.

V. FUTURE DIRECTIONS

The emergence of new language model architectures, especially GPT-4 and similar models, enable the provision of more advanced legal reasoning or even gener- ation of legal advice. Approaches that include constitutional AI might help address issues of reliability and bias by integrating legally and ethically principled approaches into the model training process from a ground-up basis [24]. Looking beyond legal reasoning, agentic systems of AI could transform access to legal services by executing autonomously routine tasks such as filing forms or generating basic legal documents. It is possible to imagine agentic systems enabling user-initiated and user-driven access to legal services for routine legal procedures that currently require professional assistance. Agentic systems could steer users through relatively standard procedures such as preparing wills and registering businesses, advancing access to legal services that are currently costly and time consum- ing efforts. There are also possibilities for enhanced accessibility using multimodal processing methods including voice input and/or image analysis of documents. Voice-based legal assistance systems could assist individuals with limited literacy or technical skills by even advancing natural language interactions, in addition to breaking down language barriers. This would also be particularly helpful for elderly users or individuals unfamiliar with written interfaces. Using optical character recognition paired with NLP (natural language processing) for scanned documents would also provide significant advancement for developing countries where users still rely on physical documents and the digital infrastructure to enable access to legal documents is limited [37]. With this capability, individuals would be able to take a photograph of a legal document and receive feedback and analysis of the document immediately without the manual requirement of inputting the document data.

By integrating with government databases and providing real-time updates on policies, legal AI systems could offer up-to-date and comprehensive information on services and procedures. For example, this would require immense coordination between AI developers and governments for data accuracy and security. Alternatively, blockchain could verify integrity and transparency of the data in such systems, particularly for more sensitive legal records, where tampering or exploitation of access poses a serious risk [38]. Secure and verifiable links to government sources would help facilitate immediate trust in the usability of AI, but would also be beneficial to ensure that the advice is current with the law or legal landscape. Explainable AI is crucial in the law, because users need a manner to understand the rationale supporting the advice to responsively interact with the AI and ascertain its correctness. Future systems should seek to relationally source the practice so that it engenders trust in the user; perhaps a citation model of the legal resource. Moreover, hybrid models of rule-based reasoning with neural nets could enhance traceability and ensure reasoning was both legitimate and understandable to laypersons [39] . Explaining reasoning of processes and citations would encourage practitioner quality assurance, which will increase transparency, and maintain public confidence in AI legal assistance.

Federated learning techniques may help to address privacy challenges, while facilitating the collaborative training of models across institutions, allowing for the use of diverse datasets to enhance legal AI systems without any data sharing that might compromise user privacy. In the context of India, for example, the sharing of anonymized data between multiple states could help to advance multilingual models [40] to ensure that systems are robust across languages and regions, while also maintaining the privacy of sensitive legal information . Approaches such as these may strike a balance between having sufficiently comprehensive training datasets and the protection of privacy, while possibly providing stronger systems overall.

New directions also include hyper-personalized AI applications and technologies to adapt solely to the individual user or user context, such as shifting language complexity through education or only offering region- specific advice while acknowledging local legal considerations.

In addition, AI ethics frameworks specific to legal tech will be vital to lessen the potential of existing risks such as deepfakes in evidence or an amplification of automated bias that would undermine justice systems. Opportunities for collaboration, such as coalitions between technology companies and legal aid or public sector organizations, may provide pathways to scale real solutions meaningfully and ethically and address other real world issues while ensuring standards of professionalism. Pilot programs in rural India for AI powered legal kiosks that provided services while also gathering data about context and the professional could help to inform the next iteration of models, based on real activity rather than exclusively relying on user reported data or datasets curated by technology teams.

VI. PROPOSED NYAYASAHAYA FRAMEWORK

NyayaSahaya is intended to be a comprehensive AI-powered service for the non-technical user in a variety of languages, especially in low legal literacy contexts like India. Previous legal AI systems have usually focused on technical professional use cases or limited legal tasks. NyayaSahaya moves past the limitations of existing legal AI to combine legal advice generation, policy simplification, and multilingual interface operation to realize one, holistic service in a multidimensional platform that can exist in a single user interface and can operate across multiple channels.

Each of the components of NyayaSahaya cooperates and interacts with each other while still executing accurate legal conclusions or disseminating simplified policy information to resource-poor users. The multilingual query processing component uses multilingual transformer models to process queries in regional languages, including a dynamic code-switching detection system that is typical for code-switching environments and multilingual societies. This capabilities allows the users to form queries in their preferred language or mix of languages naturally with each other, similar to how they would comport with each other if they were consulting a human legal advisor. The policy simplification engine uses T5-based models to generate simpler and summarize government policies that have been simplified to a specific segment of demographics to a template class of farmers or small business owners or urban workers, while still retaining the original mapping back to the legal policy, allowing for both the user's user profile and the user's combination of policy entities.

The legal advice module uses retrieval-augmented generation frameworks to supply accurate, jurisdiction-situated legal advice with guardrails against hallucination and transparent source citation. Each response includes links to relevant law, case law, or governing documents to allow users to independently confirm information or get professional help for more complicated matters. The voice and text interface accepts voice inputs and optical character recognition for accessibility, allowing users with low literacy levels to interact with the system by speaking questions or photographing documents as an alternative to typing.

Cultural contextualization takes account of socio-economic and cultural aspects, such as approaches to solving problems locally or economic limitations influencing options that simply relying on a legal recommendation might fail to account for. This frame provides explainability so that users receive not only answers, but rationale that connects the answer to a legal source, in language that is understandable to the user. Integration with government agenda will keep the advice current as policy enables real-time updates, hence advised solutions become outdated law and schemes. Pilot engagements might focus on one high impact context, for example, supporting someone to gain assistance from welfare schemes or assisting someone to understand their rights and situation in accommodation, and metrics on user satisfaction and accuracy of advice over time.. shows the layout of the system architecture, which provides a four tiers of stacks.

The user interface layer allows the user to interact with the capabilities via web, mobile, and voice sources. The NLP processing layer engages with language determination (validating English input), code-switching capabilities, and determining intent (or classification). The AI model layer employs LegalBERT for answering legal types of questions and T5 for simplifying text, and the RAG (Retrieval-Augmented Generation) model structure is utilized, which offers to validate the accuracy and reliability of any findings generated by the model. The knowledge base layer maintains repositories of legal documents, government agency polices and case law and was designed to include end=data integrations to external government API to access real-time versions of the legal policies and documents.

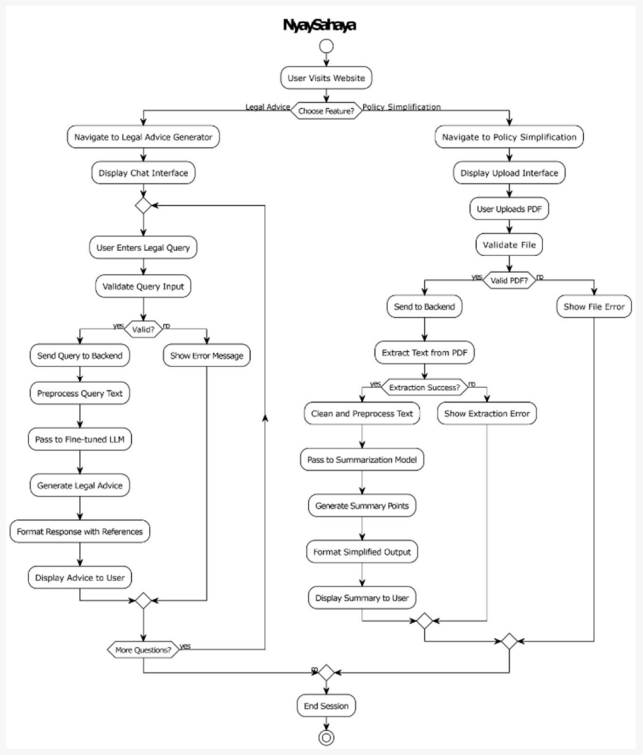


Fig. 1 System Activity Diagram

VII. CONCLUSION

The survey demonstrates considerable advancements in AI-based legal aid systems but also points to critical barriers that must be overcome if there is to be try widespread adoption or a significant positive impact on access to legal services. Current systems show great potential in areas such as legal document preparation, intelligent question answering, and legal text simplification. Yet, there are still significant restrictions in terms of reliability, multilanguage capability, and the ability to customize solutions to fit various cultures and contexts, which limit their full potential to serve different populations. Presently there are expectations that the legal AI marketplace will be valued at USD 4.9 billion by 2032 [41] .

Thus, the need for affordable, reliable accessibility is more than ever complex. The proposed system, NyayaSahaya, addresses critical gaps identified in the literature review, by integrating government policy simplification with the ability to generate legal advice, created explicitly for multilingual environments and non-technical users that do not have formal legal training. The emphasis on Indian languages and actual culture in combination with the goal of legal context makes for useful next steps towards democratizing access to legal services for the developing world, as traditional legal services are traditionally excluded to most citizens. The NyayaSahaya platform provides a scalable, easy to use platform that understands the barriers of language and socioeconomic constraints, and allows millions to have the ability to confidently and comprehend the legal worlds they exist in.

Key research directions should advocate for systems of evaluation of legal AI that are not limited to accuracy to value their operational effectiveness and the extent of user trust. The ability to create rich multilingual legal data sets will be vital as currently available data does not account for the linguistic legal richness of unique countries such as India. Ethical questions of bias and privacy will require continuous attention with the lens that legal AI can support underserved communities instead of perpetuating inequities. Legal AI systems could also influence digital governance, to strengthen citizen participation and legal compliance, governing bodies will continue advancing to deliver services digitally.

Future directions should also consider researching the effectiveness of legal AI systems in longitudinal studies in the real world to better understand their actual impact on users legal outcomes and decision making processes. The development of acceptance metrics or benchmarks for legal AI systems to evaluate benchmarks across or between systems will improve development, performance, and reliability. There should also be collaborative effort from legal AI researchers, legal professionals, policy makers, to deploy legal AI responsibly with appropriate discretion and to promote access.

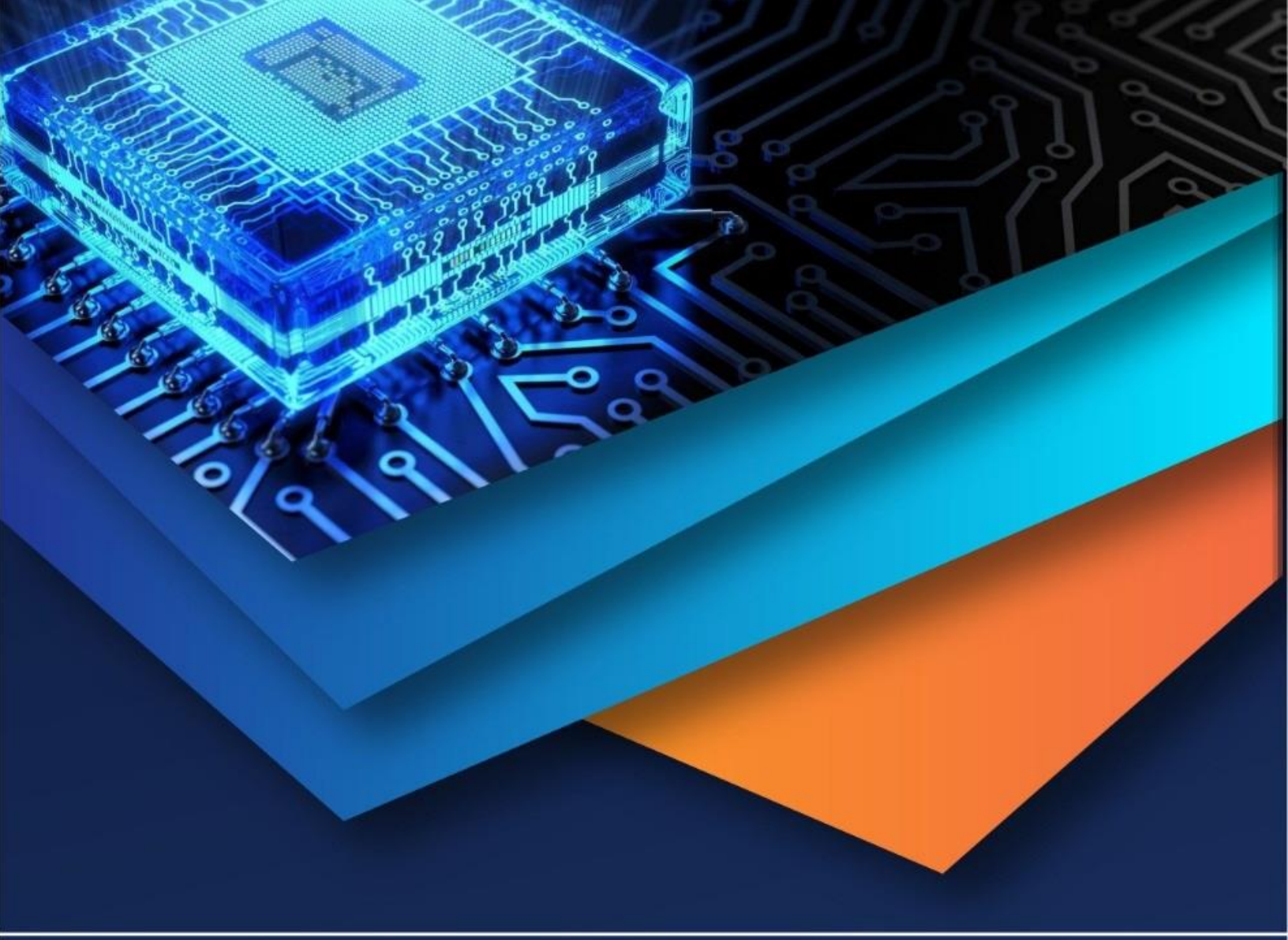
AI-driven legal access democratization presents a serious opportunity and responsibility for researchers, developers, and policymakers. Achieving this goal will necessarily involve ongoing interdisciplinary collaboration, and careful attention to and consideration of the specific needs and constraints of legal use cases, where error can have serious consequences for users. Applications such as NyayaSahaya allow us to potentially build a more just society to the extent that legal empowerment becomes available to all citizens regardless of economic circumstances, language barriers, and distance from the courts. However, in order to achieve this goal we must commit to reviewing, understanding, and measuring technological challenges, ethical concerns, and deployment barriers that are currently hindering the scalability and effectiveness of legal AI systems.

VIII. ACKNOWLEDGMENT

The authors thank legal practitioners, government representatives, and experts in the field for providing invaluable insights, constructive criticism, and commentary on this research. They express their appreciation for the community organizations in India that helped them see legal challenges at the grassroots level and understand the relevant user needs and practical constraints.

REFERENCES

[1] Singh et al., ”Legal Literacy in Developing Nations: A Comprehensive Survey,” Journal of Legal Technology, vol. 15, no. 3, pp. 245–262, 2023.
[2] M. Johnson, ”Automated Legal Services: Analysis of DoNotPay’s Impact on Consumer Legal Access,” AI & Law, vol. 30, no. 4, pp. 485–506, 2022.
[3] P. Zhang and L. Martinez, ”ROSS Intelligence: AI-Powered Legal Research Platform,” Artificial Intelligence Review, vol. 54, no. 2, pp. 1123–1145, 2021.
[4] K. Patel et al., ”Conversational AI in Legal Services: A Systematic Review,” Expert Systems with Applications, vol. 210, pp. 118–135, 2023.
[5] R. Wilson and S. Chen, ”Benchmarking AI Contract Analysis: Human vs Machine Performance,” Computer Law & Security Review, vol. 39, pp. 105–118, 2020.
[6] D. Kumar and A. Sharma, ”Simplifying Legal Documents: Approaches and Challenges,” Natural Language Engineering, vol. 25, no. 4, pp. 467–489, 2019.
[7] L. Thompson et al., ”Neural Approaches to Legal Text Simplification,” Transactions on Computational Linguistics, vol. 10,
[8] pp. 892–907, 2022.
[9] J. Lee and M. Park, ”Automatic Summarization of Legal Documents: A Survey,” Information Processing & Management, vol. 60, no. 2, pp. 103–124, 2023.
[10] N. Gupta et al., ”Government Policy Simplification Using Multi-Document Summarization,” Government Information Quar- terly, vol. 40, no. 1, pp. 45–62, 2023.
[11] F. Rodriguez and K. Tanaka, ”Multilingual Natural Language Processing in Legal Contexts,” Computational Linguistics, vol. 48, no. 2, pp. 289–318, 2022.
[12] V. Agarwal et al., ”Legal NLP for Indian Languages: Challenges and Opportunities,” ACM Transactions on Asian and Low-Resource Language Information Processing, vol. 22, no. 3, pp. 1–25, 2023.
[13] S. Sharma and P. Joshi, ”Processing Code-Switched Legal Queries in Indian Contexts,” Language Resources and Evaluation, vol. 57, no. 1, pp. 167–189, 2023.
[14] H. Wang et al., ”Named Entity Recognition in Legal Texts: A Comprehensive Survey,” Knowledge-Based Systems, vol. 245,
[15] pp. 108–127, 2022.
[16] C. Liu and D. Brown, ”Graph Neural Networks for Legal Information Extraction,” Neurocomputing, vol. 521, pp. 78–94, 2023.
[17] T. Anderson et al., ”LegalQA: A Benchmark for Legal Question Answering Systems,” Proceedings of ACL, pp. 2456–2471, 2023.
[18] I. Chalkidis et al., ”LegalBERT: The Muppets Straight Out of Law School,” Findings of EMNLP, pp. 2898–2904, 2021.
[19] G. Miller and A. Davis, ”Evaluating Large Language Models on Legal Reasoning Tasks,” AAAI Conference on AI, pp. 15234–15241, 2023.
[20] Y. Kim et al., ”T5-based Legal Document Generation and Simplification,” NAACL Proceedings, pp. 1789–1798, 2022.
[21] E. Rossi and M. Fischer, ”Cross-lingual Transfer Learning for Legal Document Classification,” EACL Proceedings, pp. 567–578, 2023.
[22] B. Taylor et al., ”Data Challenges in Legal AI: Privacy, Quality, and Availability,” AI Magazine, vol. 44, no. 1, pp. 78–92, 2023.
[23] R. Verma and S. Chopra, ”Multilingual Challenges in Indian Legal AI Systems,” ICON Conference Proceedings, pp. 245– 256, 2022.
[24] O. Nielsen and K. Jensen, ”Accuracy Requirements for AI Legal Advisory Systems,” AI & Society, vol. 38, no. 2, pp. 567–585, 2023.
[25] Q. Washington et al., ”Algorithmic Bias in Legal AI: Sources and Mitigation Strategies,” FAccT Conference, pp. 123–134, 2022.
[26] X. Chen and J. Williams, ”Constitutional AI for Legal Applications,” NeurIPS Workshop on Trustworthy AI, pp. 45–52, 2023.
[27] Thomson Reuters, ”How AI is Transforming the Legal Profession,” Thomson Reuters Legal Blog, 2025.
[28] Clio, ”AI-Powered Legal Practices Surge: Clio’s Latest Legal Trends Report,” Clio Press Release, 2024.
[29] Paxton AI, ”How AI is Transforming Legal Research and Drafting: Trends and Best Practices for 2025,” Paxton AI Blog, 2025.
[30] A. Quevedo et al., ”Legal Natural Language Processing from 2015-2022,” arXiv preprint, 2023.
[31] LexisNexis, ”Generative AI and Legal Practice,” LexisNexis Survey, 2024.
[32] S. Kumar et al., ”Natural Language Processing for Legal Documentation in Indian Languages,” ResearchGate, 2024.
[33] A. Kanapala et al., ”Indian Legal NLP Benchmarks: A Survey,” arXiv preprint, 2021.
[34] IndiaAI, ”Advancing Multilingual NLP for Indian Legal Systems,” IndiaAI Report, 2024.
[35] G. Miller et al., ”AI-Powered Lawyering: AI Reasoning Models,” SSRN, 2025.
[36] Stanford HAI, ”AI on Trial: Legal Models Hallucinate in 1 out of 6 Queries,” Stanford HAI News, 2024.
[37] Thomson Reuters, ”Legal Issues with AI: Ethics, Risks, and Policy,” Thomson Reuters Blog, 2025.
[38] American Bar Association, ”Big Data, Big Problems: The Legal Challenges of AI-Driven Data,” ABA Business Law Today, 2024.
[39] IE University, ”The Future of AI in Law: Trends and Innovations to Watch,” IE Uncover, 2025.
[40] ContractPodAi, ”Future Legal Departments in 2030: AI & Automation,” ContractPodAi News, 2025.
[41] McKinsey, ”AI in the Workplace: A Report for 2025,” McKinsey Report, 2025.
[42] Google Research, ”Federated Learning for Multilingual NLP,” Google Research Blog, 2024.
[43] Yahoo Finance, ”Legal AI Market Size to Surpass USD 4.9 Billion by 2032,” Finance Report, 2025.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)