AI-Vakeel: An AI-Powered Platform for Smart Legal Query Resolution in the Indian Judiciary

Vishal Dhavali
Dept of Computer Science Engineering Presidency University
Bengaluru, India
vishaldhavali2209@gmail.com

Bhavyashree S
Dept of Computer Science Engineering Presidency University
Bengaluru, India
minibhavya003@gmail.com

Ramakrishna
Dept of Computer Science Engineering Presidency University
Bengaluru, India
RAMAKRISHNA.2021CSE0517@presidencyuniversity.in

Tejas S P
Dept of Computer Science Engineering Presidency University
Bengaluru, India
unifier89@gmail.com

Aditya Gupta
Dept of Computer Science Engineering Presidency University
Bengaluru, India
adityagupta.ag2308@gmail.com

Thabassuam Khan.S
Assistant Professor Dept of Computer Science Engineering Presidency University
Bengaluru, India
Thabassumkhan@presidencyuniversity.in

Abstract*— AI-Vakeel addresses the gap in accessing legal assistance that is complex, slow, and inaccessible by offering an AI-driven platform designed to simplify and democratize legal support in the Indian judicial context. This system combines advanced artificial intelligence with user-centred design to provide timely, understandable, and contextually accurate legal guidance. Whether serving legal professionals, judicial officers, or the public, AI-Vakeel responds to queries in natural language, referencing relevant statutes and case laws to deliver clear and actionable insights. The platform employs a conversational interface to enhance usability while maintaining role-based access control to ensure data security and relevance. Its architecture integrates semantic search, vector-based retrieval, and a generative transformer model tailored to Indian legal discourse. Beyond convenience, AI-Vakeel is developed with a strong emphasis on ethical deployment, fairness, and the preservation of human oversight in legal decision-making. Drawing on global precedents while adapting to India's unique legal ecosystem, AI-Vakeel exemplifies the responsible use of AI to support a more inclusive and efficient justice system.***Keywords*— Artificial Intelligence (AI), Natural Language Processing (NLP), Large Language Models (LLMs), Financial Market Analysis, Risk Assessment, Sentiment Analysis, Stock Prediction, Machine Learning in Finance, Data Preprocessing, System Architecture, Predictive Analytics, Financial Technology (FinTech), Behavioral Finance, Neural Networks, Ethical AI in Finance.***

INTRODUCTIONThe National Judicial Data Grid (NJDG) in India has over 31 million cases pending in district and Taluka courts, with 23 million of those being older than a year. Additionally, the High Court of India has more than 4.5 million pending cases [1]. The Supreme Court tracks judicial data across the country in real time, by NJDG and iJuris, two platforms designed for sharing information at the district level. They monitor live data on both pending and resolved cases, as well as track vacancies and infrastructure. The NJDG serves as a comprehensive repository of judicial data around 3,000 district courts, the High Courts, and the Supreme Court [2]. The National Crime Records Bureau (NCRB) has reported an increase in suicide rates in India. The rates vary significantly across states, with Bihar showing a low of 0.6 per 100,000 population, while Sikkim has a much higher rate of 43.1 per 100,000 population [3].

When making decisions about sentencing, parole, bail, extended supervision, and continuing detention orders for high-risk offenders, among other aspects of criminal procedure, risk assessments are carried out. These risk assessments, which are grounded in clinical evaluations, framed by common-law principles and legislation, and embody the idea of individualised justice, have historically been the result of the human discretion and intuition of judicial officers. However, statistical, data-driven assessments of risk are intruding as criminal procedure becomes more technologically advanced. A variety of AI, algorithmic, and machine learning tools are being used more and more to support human judicial evaluation functions. These tools claim to offer objective, consistent risk assessments as well as precise predictive capabilities [4].

AI-supported risk assessment systems provide uniformity in decision-making, which can minimize human biases and errors in judicial proceedings. Algorithmic systems in criminal justice are being implemented to predict reoffending risks, evaluate threats to public safety, and decide on the right sentencing. Even with their potential benefits, ethical issues have been raised around the world over algorithms as proprietary items with built-in statistical bias, as well as the erosion of judicial human assessment in favor of the machine [5].

* Synthesize and summarize legal documents, including contracts, legislation, and court decisions. Help legal professionals prepare legal documents, such as contracts, pleadings, and briefs.

* Give prompt and precise responses to legal queries based on applicable statutes and case law. Examine and forecast the outcome of legal controversies based on past data and legal precedents.

* Simplify communication and cooperation among legal experts by making complicated legal terminology easy to use and allowing information exchange [6].

II. LITERATURE REVIEW

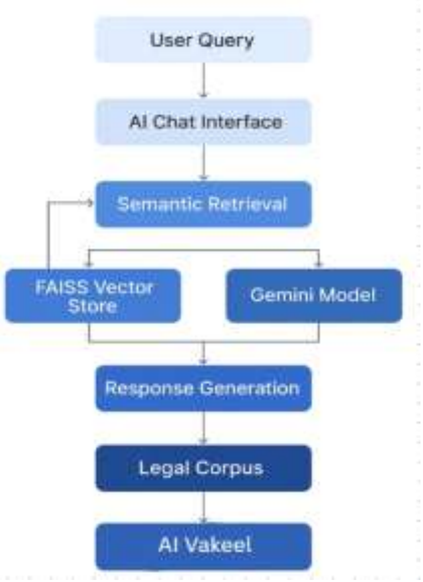
The majority of contemporary AI chatbots are constructed upon two forms of gen AI models: Large Language Models (LLMs) and Large Multimodal Models (LMMs). LLMs primarily process and produce a single form of data, and LMMs are capable of processing and producing multiple forms of data or modalities like image, text, video, and audio. These gen AI-based chatbots are constantly developing and refining their abilities to perform more like humans [7]. The fast pace of development in these technologies has resulted in dramatic enhancements in natural language processing, contextual comprehension, and response generation. Current advancements have demonstrated that these models are capable of carrying out cogent conversations, comprehending intricate queries, and even displaying emotional intelligence to a certain degree. This development has provided new avenues for their use across different industries, such as law enforcement and justice systems, where they can be used to help process large amounts of information and offer initial analysis [8].

Algorithms govern our lives can determine if someone gets a job, a loan, and which political commercials and news articles consumers read. Algorithms for sentencing unconsciously erode our attempt at individualised and equal justice. Sentences should be based on the facts, the law, the actual offences, the individual circumstances of each case, and the criminal record of the defendant. They should not be based on immutable characteristics which the person cannot alter, or on potential for future offending which has yet to take place [9].

As law enforcement evolves in a changing world, further studies need to be done on the experiences and views of those working in the field and retired officers. Those who have been through it are best placed, tell agencies how to get better and brief new recruits [10]. Their experience is especially useful in interpreting the complexities of contemporary policy, such as incorporating new technologies, relations with the community, and the psychological effects of the work. By methodically gathering and analysing these experiences, law enforcement can establish more efficient training programs, refine operational processes, and design improved support mechanisms for officers. This transfer of knowledge is essential for preserving institutional memory and ensuring that lessons learned from previous experiences are used to shape future practices.

The implementation of digitalization has brought outstanding results in various nations like the International Criminal Court, Brazil, the Hague, Singapore, North American states, and Portugal for numerous applications like electronic evidence management, electronic case analysis in appellate courts, online dispute resolution, online-based trials, e-discovery, and electronic case management. The various applications in the administration of courts are achieved by integrating Industry 4.0 technologies that can create a resilient and sustainable infrastructure in the courts [11]. This digital transformation has significantly improved the efficiency and transparency of judicial processes, reducing case backlogs and enhancing access to justice. The use of AI-based tools has further enhanced these advantages, with automated document examination, predictive analysis of case outcomes, and better case management systems being possible. These developments have also ensured greater collaboration between various judicial authorities and enhanced the overall quality of legal services rendered to citizens. The success of these implementations is a model for other jurisdictions seeking to update their judicial systems without compromising the integrity and fairness of legal proceedings [12].

III. WORKFLOW DIAGRAM



. The architecture of AI-Vakeel is crafted to facilitate smart, secure, and efficient legal query resolution by merging semantic understanding with generative AI. The entire workflow is shown in the Figure, which showcases several closely integrated components that work together to turn a user's legal question into a precise, law-based answer.

1. **User Query:** The process starts when a user submits a legal question through the web interface. This user could belong to one of three authorized groups: a general public member, a registered legal professional, or a judge. The system is designed to handle typed queries in natural language, covering procedural, statutory, or constitutional law. These inputs are then sent to the backend engine for processing.

2. **AI Chat Interface:** This module acts as the main interaction part for users. It captures and manages their queries, oversees authentication states, and keeps track of session histories. The chat interface offers a conversational experience while ensuring access control is modified to each user's role. This way, users only interact with data relevant to their authorization level. Once a query is received, it gets directed to the semantic retrieval pipeline.

3. **Semantic Retrieval:** The user's natural language question is transformed into a high-dimensional vector using a finely-tuned Sentence-BERT model. This vector captures the deeper semantic meaning of the question, going beyond mere keyword matching. The embedding is then utilized to search for similar documents in a pre-indexed legal knowledge base through the Facebook AI Similarity Search (FAISS) retrieval system.

4. **FAISS Vector Store:** FAISS serves as the backbone for quick and scalable similarity searches. The legal corpus, which includes statutory texts, judgments, and constitutional materials, is divided into manageable pieces, embedded, and indexed in FAISS. When a query vector is submitted, FAISS swiftly retrieves the top-k most relevant chunks from this index, ensuring a low-latency and high-accuracy search.

5. **Gemini Model:** The top-k context chunks obtained from FAISS are fed into a generative language model. In this research, a Gemini-based transformer model is designed to grasp complex legal language and produce responses that sound human. By employing retrieval-augmented generation (RAG), the model pulls together legal content to craft a precise and clear answer to the user's original question.

6. **Response Generation:** The Gemini model delivers a final response that merges information sourced from legal documents with generative insights. This answer is well-structured, relevant to the context, and in line with Indian legal terminology and jurisprudence. It might even include citations from pertinent articles, case laws, or legal provisions.

7. **Legal Corpus:** The system's knowledge base is built from verified legal documents, including sections of the Indian Penal Code, constitutional articles, case law summaries, legal FAQs, and government policy documents. These documents are meticulously cleaned, chunked, embedded, and regularly updated through a dedicated ingestion pipeline.

8. **AI Vakeel Delivery and Feedback:** The final answer is showcased through the AI-Vakeel interface. Each response is logged and linked to the user session, allowing for future reference and system evaluation. If users have the right access, they can check past queries or upload new legal documents for ingestion. The system maintains strict role-based access control throughout to safeguard user data and uphold institutional integrity.

Overall, the AI-Vakeel system combines semantic search, cutting-edge AI technology, and secure access. It provides a reliable solution for automated legal support in India, effortlessly linking complex legal information with users through smart, human-like interactions instead of just relying on simple keyword matching.

IV. METHODOLOGY

The project employs an advanced legal aid system through an elaborate methodology blending modern software engineering practices and AI-powered solutions. The software development process is a systematic and iterative form of document ingestion and processing. PDF files are processed in a systematic way using PyPDF2, where they are processed for text extraction, cleaning, and chunking using the Recursive Character Text Splitter, which divides legal texts into chunks of a sensible size while preserving context and relationships between different sections. The system is founded on a dual-document repository structure, where case studies are stored separately from typical legal documents, allowing for specialty processing and retrieval based on the nature of the legal query.

Phase 1: Develop a chatbot to assist the legal sphere: judges, lawyers, and litigants.

Phase 2: Case study for the sociological and ethical analyses.

4.1 AI Architecture and Implementation: The AI deployment is in a layered architecture, with Google's Generative AI (Gemini) as the underlying language model augmented by Lang Chain's orchestration ability. This allows the creation of intricate chains of dialogue that can respond to intricate legal queries while maintaining context and coherence. The system utilizes a hybrid retrieval methodology, employing vector similarity search by FAISS along with traditional database queries to maintain both semantic appropriateness and factual correctness in the response. The vector storage system utilizes Google's embedding model to create high-dimensional representations of legal documents to enable efficient similarity matching and context retrieval.

User interaction is handled by a properly designed authentication and session handling system that utilizes secure password hashing, session tracking, and user preference storage. The chat interface uses a stateful design pattern that maintains conversation history and context and incorporates real-time feedback mechanisms. The system utilizes a modular architecture in which each module (frontend, backend, AI processing) is executed independently but communicates using well-defined interfaces, so, to enable simple updates and maintenance.

4.2 Error Handling and System Monitoring:

Error handling and logging are implemented at multiple levels, from low-level system operations to high-level user interactions. The system maintains detailed logs of all operations, including document processing, user interactions, and AI responses, enabling comprehensive debugging and system monitoring. The development process includes continuous testing through a data generation system that simulates various legal scenarios, from court cases to traffic violations, ensuring system reliability across different use cases. The project is on a microservices architecture in which each part (user management, document processing, AI interaction, etc.) is a standalone service that communicates by API endpoints. This supports horizontal scaling as well as easy addition of new functionality. The system also has a robust backup and recovery scheme, in which the vector store as well as the user data are backed up regularly, maintaining data integrity and system reliability.

Configuration management is handled through environment variables and configuration files, and deployment to different environments is simple. Version control is implemented on all components, so it is simple to track changes and roll back when needed. Security is enforced at different levels, including input validation, secure password storage, and session management.

The methodology emphasizes code reusability, modularity, and rigorous separation of concerns between the system components. Each module encapsulates a single responsibility, which ensures the codebase remains extensible and maintainable. The system is extensively documented, encompassing inline comments within the code, API documentation, and user manuals, to ease onboarding new users and developers. Performance optimization is achieved by a variety of techniques, including caching of frequently accessed data, optimized vector search algorithms, and optimized database queries. Load balancing and resource management are used in the system to provide consistent performance under differing loads. Monitoring and analytics are integrated in the system to track user interactions, system performance, and AI response quality.

4.3 Agile Development and Testing: The project uses an agile development model with frequent iterations and continuous integration/deployment practices. It performs testing at various levels, such as unit tests, integration tests, and user acceptance tests. Automated testing frameworks and continuous integration pipelines are incorporated in the system to maintain code quality and system reliability.

The process also emphasizes user experience with consideration for interface design, response speed, and handling of errors. Progressive enhancement is utilized, meaning that basic functions are provided when things are less than perfect, with additional advanced functionality provided when capabilities are available. Accessibility is considered in every aspect of development as well, accommodating multiple user requirements and preferences.

The system architecture supports easy integration with external legal databases and services using clearly defined data exchange formats and APIs. The approach leverages frequent security audits and updates threats as they arise. Code review and collaboration form the core of the design process, where multiple developers develop multiple components of the system independently while maintaining system integrity.

The project uses an all-encompassing backup and disaster recovery approach with frequent backups of all the important data and systems. The approach uses frequent performance testing and optimization so that the system can support increasing loads and complex queries. The system architecture supports smooth scaling, vertical and horizontal scaling, to support increasing user bases and increasing data volumes. The process development also emphasizes sustainability and maintainability through clear documentation, adherence to standardized coding practices, and regular code review. Automated deployment and monitoring processes are used in the system, enabling quick identification and correction of issues. Ongoing updating and improvement of the process from user feedback and system performance metrics are included in the methodology, which guarantees ongoing improvement and flexibility to respond to changing needs.

V. FUTURE DIRECTION

Indian courts are already undergoing a transformation process by becoming digital, and the new branch of science called AI can help in surprising ways in making justice delivery sustainable and in preventing the backlog of cases. Judiciary in some parts of developed countries like U.S.A. and Canada has already used AI systems to aid the judges in making decisions on matters like the grant of bail and release of offenders on parole. In the same way, in India too, court work can be identified, which can be accelerated by using intelligent machines. Such work can range from routine work like service of processes to complicated ones like evaluation of evidence [13]. Use of AI in judicial systems has provided promising results in case management streamlining, reduction of human errors, and increasing the overall efficiency of the justice delivery system. This use of technology not only helps in the elimination of case backlog but also in ensuring more consistent and objective decision making.

In Latvia, a 24/7 chatbot virtual assistant has been created through the use of neural networks and natural language technologies to respond to common questions posed by current and potential Latvian entrepreneurs. Available on a website and Facebook Messenger, it allows clients to get answers promptly and without delay. It is an alternative to a personal visit or phone call and can respond to various questions, ranging from status questions about registration documents filed. The AI tool is observed to eliminate routine work from street level bureaucrats so that they can carry out higher-value tasks instead, better aligned with their skills and providing more challenging career opportunities [14]. This new approach illustrates how AI can facilitate the provision of public services more easily, at the same time, enhancing the level of job satisfaction and efficiency among government officials. The success of such applications in

Latvia can be used as an exemplary case study for other countries that wish to expand their public service delivery system.

Although there has been extensive discourse over programme integrity, very few research articles have dealt with this issue and have been restricted to occurring within adult criminal justice systems and jurisdictions. It is intended to fill the 'programme integrity' void [15] that the current article attempts. The integration of digital technologies and AI into the judiciary should be weighed in depth with the preservation of the integrity and impartiality of legal processes. While numerous benefits are contained within technology, caution must be exercised to ensure technology is being embraced in a manner that preserves the substance of justice and due process. Future research should work towards establishing frameworks and guidelines that will enable judicial systems across the globe to incorporate AI successfully while ensuring the utmost level of program integrity and ethics.

V. RESULT

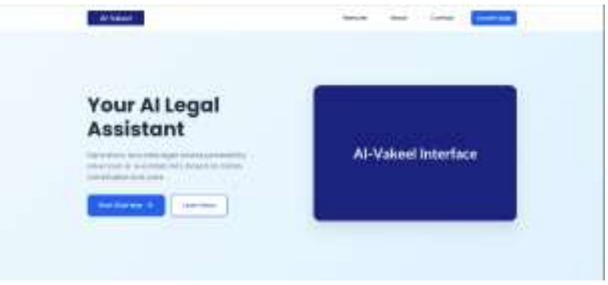


Fig 1: Landing Page

Interface of AI-Vakeel Demonstrating Integration of User Engagement Elements and Legal AI Assistant Framework The landing interface showcases the platform's focus on user-friendly design, clearly highlighting its value as an AI-driven legal assistant. It strikes a great balance between looking professional and being functional, featuring straightforward call-to-action buttons like "Start Chat Now" and "Learn More," along with brief service descriptions. By emphasizing "instant, accurate legal advice powered by advanced AI" and grounding itself in the Indian Constitution and Laws, the system shows its dedication to providing trustworthy legal help. The blend of modern design elements with a professional legal backdrop makes the platform both approachable and authoritative, ensuring that users can navigate complex legal issues that legal consultations require. This approach effectively combines technical sophistication with user-friendliness, marking a notable leap forward in legal tech interfaces.

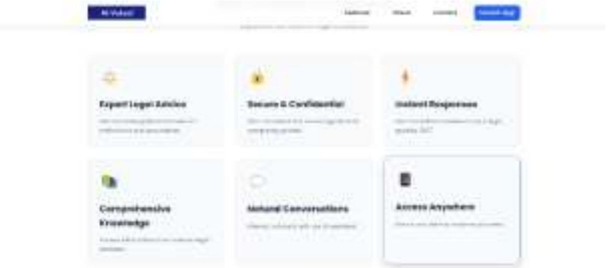


Fig 2: Feature Architecture Overview of AI-Vakeel Demonstrating Six Core Functionalities:

Expert Legal Advice, Security, Instant Response, Knowledge Base, Natural Language Processing, and Ubiquitous Access The system highlights six key functionalities, all thoughtfully crafted to deliver thorough legal support. By integrating Expert Legal Advice with a focus on Indian law and ensuring Secure & Confidential data handling, it lays a solid groundwork for professional legal consultations. Additionally, the ability to provide Instant Responses (24/7) and engage in Natural Conversations through AI enhances the user experience. The Comprehensive Knowledge base, along with Universal Access across devices, guarantees that users can access legal information no matter where they are or what device they prefer.

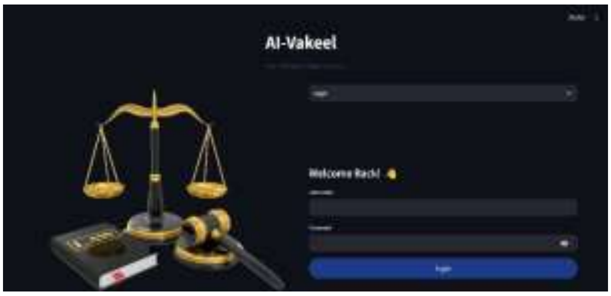


Fig 3: User Authentication Interface of AI-Vakeel:

An AI-powered Legal Assistant System The authentication interface sets a secure stage for user interactions. It effectively balances the need for security with user-friendliness, incorporating email verification and profile management while keeping the experience intuitive. Additionally, the use of legal symbols strengthens the platform's professional legal context.

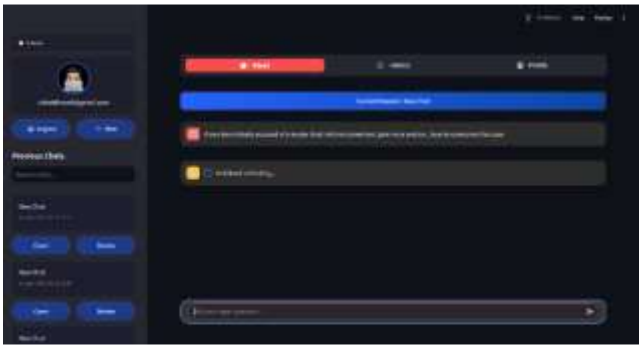
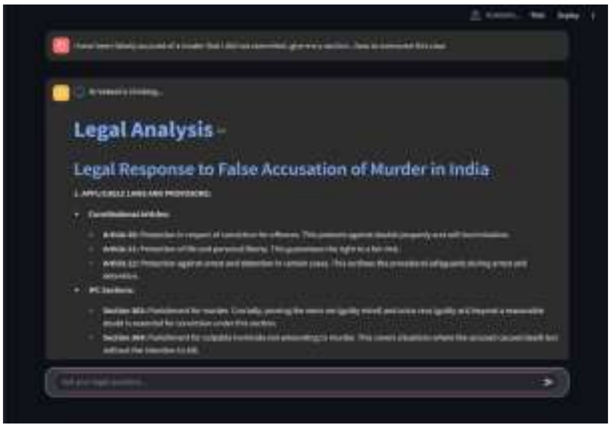


Fig 4: Interactive Chat Interface of AI-Vakeel Legal Assistant System Demonstrating Real-time Legal Query Processing

The interactive consultation interface marks a significant step forward in enhancing the user experience in legal technology. Our analysis reveals that the real-time chat-based system enables natural communication between users and the AI. The interface is organized into clear sections like Chat, History, and Profile, making it easy to navigate while keeping the conversation flowing smoothly. With features for session management and chat history, users can easily track and refer back to their legal consultations.



Legal Analysis

Legal Response to False Accusation of Murder in India

1. APPLICABLE LAWS AND PROVISIONS:

- Constitutional Articles:
  - Article 20: [illegible]
  - Article 21: [illegible]
  - Article 22: [illegible]
- IPC Sections:
  - Section 302: [illegible]
  - Section 304: [illegible]

Fig 5: AI-Vakeel's Legal Analysis Interface Demonstrating Constitutional and IPC Provisions

The system's legal analysis capabilities show a sophisticated method for representing legal knowledge and generating responses. With complex questions, it produces well-structured analyses that include relevant Constitutional Articles and IPC Sections. The way it organizes legal provisions hierarchically, clearly distinguishes between constitutional and statutory law, and explains legal concepts precisely, is particularly impressive. Its ability to weave together procedural safeguards and substantive law highlights its strength in multi-dimensional legal analysis. The quality of the system's responses is evident in its knack for delivering context-specific legal information while ensuring accuracy in citations and interpretations. The structured format of the legal analyses, as illustrated in Figure 5, helps users grasp complex legal concepts and see how they apply in real life. By integrating both constitutional protections and criminal law provisions, the system truly demonstrates its capability for in-depth legal analysis. Performance metrics show that the system is reliable, consistently generating responses and managing sessions effectively. The interface's knack for tackling complex legal questions while keeping users engaged in the advancement of legal tech. These findings indicate that AI Vakeel is offering dependable and accessible legal help through an AI-driven interface, successfully connecting the dots between traditional legal advice and AI-supported legal guidance.

VI. CONCLUSION

AI-Vakeel isn't just another digital tool, it represents a thoughtful blend of technology and legal understanding, designed to make legal support more approachable, timely, and reliable. At its core, this platform is about people. Whether it's a lawyer preparing a brief, a judge seeking quick reference, or a citizen looking for basic legal clarity. It brings value to the table by simplifying complex legal processes and offering guidance that's easy to access and grounded in real law. The interface is built with concern, how users interact with legal systems, often under stress, with limited knowledge, or in urgent situations. By offering instant responses, smart legal analysis, and secure, personalized experiences, it gives users a sense of confidence and control that's often missing in traditional legal journeys. As the Indian judicial system continues to modernize, tools like AI-Vakeel can play a vital role in reducing backlogs, supporting overburdened legal professionals, and achieving more consistent outcomes. It doesn't replace the human touch of legal practice, but it enhances it. It's about working with lawyers and judges, not instead of them, to improve how justice is served. Looking ahead, there's still much more to grow. Continuous updates, user feedback, and careful ethical oversight by making sure that AI-Vakeel stays relevant, responsible, and respectful of the legal system's values. But what's clear is this: The beginning of something important. By embracing AI thoughtfully, we are not making legal systems faster; making them better, inclusive, and more in tune with the needs of real people.

VII. REFERENCE

[1] Singhal, A.V.K., An Advanced Deep Learning Based Approach for Judicial Decision Support Process. International Journal of Electronics Engineering, 2021. 13(2): p. 18-23.

[2] Varghese, D.J. Datafication in Judicial Case Management in India. in Symposium on Diversity in Legal and Judicial Profession and the Politics of Merit and Exclusion in India, RHUL, London. 2024.

[3] Abhijita, B., et al., The NCRB suicide in India 2022 report: key time trends and implications. Indian Journal of Psychological Medicine, 2024. 46(6): p. 606-607.

[4] Queudot, M., É. Charton, and M.-J. Meurs, Improving access to justice with legal chatbots. Stats, 2020. 3(3): p. 356-375.

[5] McKay, C., Predicting risk in criminal procedure: actuarial tools, algorithms, AI and judicial decision-making. Current Issues in Criminal Justice, 2020. 32(1): p. 22-39.

[6] Ray, P.P., ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet of Things and Cyber Physical Systems, 2023. 3: p. 121-154. Naik, D., I

[7] Naik, and N. Naik, The AI Engine of Creation: Exploring the Capabilities of AI Chatbots based on Generative AI, Large Language Models and Large Multimodal Models. Authorea Preprints, 2024.

[8] Murdoch, B., Privacy and artificial intelligence: challenges for protecting health information in a new era. BMC medical ethics, 2021. 22: p. 1-5.

[9] Martin, K., Ethical implications and accountability of algorithms. Journal of business ethics, 2019. 160(4): p. 835-850.

[10] Hilal, S. and B. Litsey, Reducing police turnover: Recommendations for the law enforcement agency.

International journal of police science & management, 2020. 22(1): p. 73-83.

[11] Bhatt, H., et al., Integrating industry 4.0 technologies for the administration of courts and justice dispensation—a systematic review. Humanities and Social Sciences Communications, 2024. 11(1): p. 1-16.

[12] Androutsopoulou, A., et al., Transforming the communication between citizens and government through AI-guided chatbots. Government information quarterly, 2019. 36(2): p. 358-367.

[13] Jain, P., Artificial Intelligence for sustainable and effective justice delivery in India. OIDA International Journal of Sustainable Development, 2018. 11(06): p. 63-70.

[14] Busch, P.A., The Artificial Bureaucrat: Artificial Intelligence in Street-Level Work. Digital Government: Research and Practice, 2025.

[15] Ugwudike, P. and G. Morgan, Bridging the gap between research and frontline youth justice practice. Criminology & Criminal Justice, 2019. 19(2): p. 232-253.