

# PraHar AI Legal Assistant

Pratik Pramod Harugade

Department of Computer Science and Information  
Technology, Chhatrapati Shivaji Maharaj University, Navi  
Mumbai, India

Dr. Vikas Kumar

Professor & Head, Department of Computer Science and  
Information Technology, Chhatrapati Shivaji Maharaj  
University, Navi Mumbai, India

**Abstract** - The manual process of legal research and document drafting is often characterized by information overload, high costs, and potential human error. This paper introduces PraHar AI, a domain-specific multi-agent legal assistant built on the CrewAI framework. Unlike general-purpose chatbots, PraHar AI utilizes a collaborative ecosystem of specialized AI agents to automate legal issue classification, retrieve statutes and precedents, and draft structured legal documents. By integrating Retrieval-Augmented Generation (RAG) and Llama 3.3 (70B) through the Groq LPU inference engine, the system ensures high-speed processing with factual grounding. Experimental evaluations demonstrate that this multi-agent approach significantly enhances consistency and efficiency compared to single-agent systems, particularly in the Indian legal context.

**Keywords** - AI Legal Assistant, CrewAI, Multi-Agent System (MAS), Retrieval-Augmented Generation (RAG), Legal Automation, NLP.

## I. INTRODUCTION

The legal profession involves complex tasks such as case law analysis, research into statutes, and drafting petitions or contracts. Lawyers and students frequently spend excessive time on monotonous tasks that are prone to inefficiencies. **PraHar AI** aims to automate these workflows by leveraging multiple AI agents that collaborate in a structured reasoning pipeline. The system provides an end-to-end solution for issue classification, research, and documentation. PraHar: AI Legal Assistant uses CrewAI, a framework that enables several AI agents to collaborate, to automate these procedures. PraHar provides an integrated solution that handles issue classification, legal research, and document drafting, in contrast to single-task AI tools.

## II. LITERATURE REVIEW

Recent studies have highlighted the transformative role of AI in legal decision-making and analytics. While tools like DoNotPay provide automated advice for simple cases, and RAG systems assist in case law retrieval, several research gaps remain. These include limited multilingual support, a lack of end-to-end automation, and weak jurisdiction-specific adaptation. PraHar AI addresses these gaps by utilizing CrewAI for orchestrated reasoning and Streamlit for a user-friendly multilingual interface.

## III. RESEARCH GAPS AND MOTIVATION

- Based on the literature review of existing legal AI systems, several critical gaps have been identified which motivate the development of PraHar AI:
- Lack of End-to-End Automation: Most current tools

are fragmented and only handle specific tasks like search or drafting, but fail to provide a complete pipeline from query to report.

- Limited Multilingual Support: There is a significant scarcity of legal AI assistants that can accurately process and respond in regional Indian languages like Marathi and Hindi.
- Weak Jurisdiction-Specific Adaptation: General-purpose LLMs often provide generic legal advice that does not align with specific Indian statutes like the IPC or the Indian Contract Act.
- Poor Explainability and Transparency: Existing systems often function as "black boxes," failing to provide clear citations or legal reasoning for their outputs.
- Minimal Integration with Live Legal Databases: Current systems often rely on static, pre-trained data rather than real-time integration with databases like Indian Kanoon or SCC Online.
- Lack of Real-time Legal Query Handling: High latency in processing complex legal reasoning makes most systems impractical for real-time consultation.

## IV. HOW PRAHAR AI BRIDGES THESE GAPS

- **End-to-End Workflow:** By using CrewAI, the system automates the entire journey from issue classification to generating a professional Word report.
- **Native Multilingual Support:** The integration of the langdetect library and specialized prompting allows PraHar AI to handle queries in Marathi, Hindi, and other regional languages flawlessly.
- **Contextual Grounding (RAG):** The system uses Retrieval-Augmented Generation to ensure that every response is strictly based on the uploaded document, eliminating hallucinations.
- **Explainable AI (XAI):** Each summary produced by the agents includes specific clause references and risk categories, ensuring full transparency for the user.
- **High-Speed Reasoning:** Leveraging Groq's LPU allows the system to perform complex legal auditing in sub-second timeframes (\$0.8-1.5\$ seconds).

## V. PROBLEM STATEMENT

The manual process of legal research and document auditing in India is plagued by significant challenges, including information overload, high operational costs, and a heavy reliance on human expertise. These traditional methods are not only time-consuming but also highly susceptible to human error, which can compromise the reliability of legal outputs. Furthermore, a vast majority of the population faces a "Justice Gap" due to the complexity of "legalese" and the lack of advanced, high-reasoning AI tools that support regional languages like Marathi and Hindi. Current general-purpose AI systems often suffer from "hallucinations," lack factual grounding (RAG), and fail to provide explainable legal citations or structured reporting. Consequently, there is an urgent need for an integrated, multi-agent AI framework like **PraHar AI** that can automate document auditing, provide risk categorization, and ensure linguistic inclusion through real-time, multilingual legal assistance.

## VI. OBJECTIVES

The primary objective of this research is to develop **PraHar AI**, a robust multi-agent framework designed to automate and streamline complex legal workflows. The system aims to achieve automated legal query classification and issue identification using advanced natural language understanding. A core goal is to enable efficient legal research and document auditing by leveraging **Retrieval-Augmented Generation (RAG)** to retrieve relevant statutes, precedents, and IPC sections from unstructured PDF files. Furthermore, the project focuses on supporting multilingual queries in regional languages like Marathi and Hindi to ensure legal inclusivity for a broader demographic. Additionally, the framework is designed to automate the drafting of structured legal contracts and petitions, providing a risk assessment engine that categorizes obligations based on severity. Finally, the research seeks to provide a user-friendly interface for generating downloadable, color-coded professional Word reports, thereby reducing the manual effort and time required in traditional legal research processes.

## VII. PROPOSED SYSTEM

**PraHar: AI Legal Assistant** The proposed framework, **PraHar AI**, is engineered as a domain-specific legal assistant utilizing a multi-agent orchestration pattern via the **CrewAI** framework. Unlike conventional single-prompt AI systems, PraHar AI employs a collaborative workforce of specialized autonomous agents to ensure high-fidelity legal auditing. The architecture integrates an **Issue Classifier** to categorize legal queries and a **Legal Researcher** that performs semantic retrieval of statutes and precedents from unstructured PDF documents using **Retrieval-Augmented Generation (RAG)**. A dedicated **Document Drafter** and **Response Generator** synthesize the extracted information into structured legal drafts and simplified summaries. The core reasoning is powered by the **Llama-3.3-70b-versatile** model, executed through **Groq's Language Processing Units (LPUs)** to achieve sub-second inference latency. Furthermore, the system includes a multilingual logic module that automatically identifies regional languages like Marathi and Hindi, ensuring

linguistic inclusivity. The final output is rendered through a **Streamlit** interface, which also facilitates the generation of downloadable, color-coded professional Word reports with automated risk assessment.

## VIII. ARCHITECTURE

The architecture of PraHar AI is designed as a decentralized, multi-agent ecosystem optimized for high-reasoning legal tasks. The workflow begins with the User Input layer, where legal issue details are ingested via a Streamlit-based interface. This input is then passed to the CrewAI Framework, which acts as the primary orchestration engine. The framework manages a sequential pipeline of autonomous agents, each assigned specific roles and tasks:

- **Case Intake Agent:** Facilitates the initial data ingestion and identifies the core legal parameters of the query.
- **IPC Section Agent:** Utilizes a specialized **RAG-based search tool** to retrieve relevant sections from the Indian Penal Code (IPC) and other statutes.
- **Legal Precedent Agent:** Employs the **Tavily Search tool** to perform real-time retrieval of relevant precedent cases and judicial rulings.
- **Legal Document Drafting Agent:** Synthesizes the outputs from the preceding agents to draft structured legal documents, ensuring factual grounding and professional formatting.

The final output, consisting of the synthesized legal information or drafted documents, is rendered back to the user through the **Streamlit UI**, completing the end-to-end automated cycle. This architecture ensures that each stage of the legal auditing process is cross-verified by specialized agents, thereby minimizing hallucinations and enhancing accuracy.

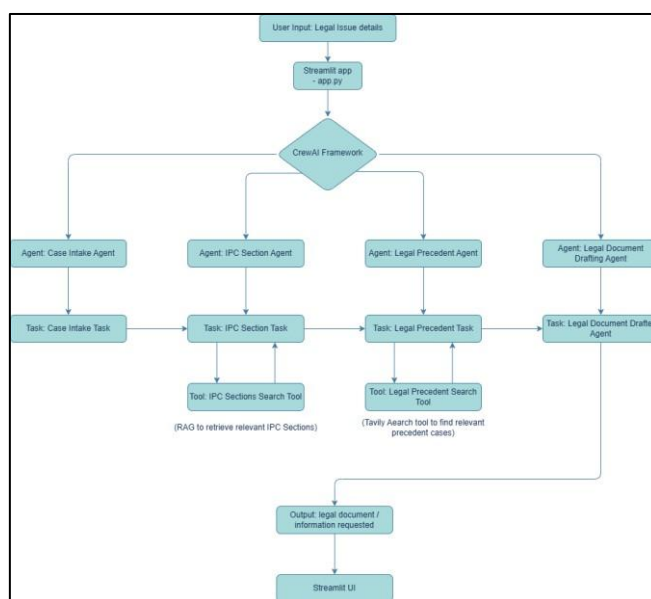


Fig. 1 shows the overall system architecture.

## IX. IMPLEMENTATION

- The implementation of **PraHar AI** leverages a state-

of-the-art Python-based stack, integrating agentic orchestration with rapid inference capabilities. The system is architected as follows:

- **Frontend and Interface:** The user interface is developed using **Streamlit**, facilitating a seamless PDF upload mechanism and a multilingual query input field. The interface manages real-time status updates and provides a trigger for the multi-agent analysis.
- **Document Ingestion and Parsing:** For data extraction, the system utilizes **LangChain's PyPDFLoader**. To ensure contextual relevance while staying within token limits, the document is split into chunks, and the primary context is derived from the first eight pages of the legal PDF.
- **Agentic Orchestration Framework:** The core reasoning is managed by **CrewAI**. Two specialized agents are instantiated:
- **Senior Legal Researcher:** Configured with a goal to extract specific clauses and penalties based on the user's query.
- **Legal Drafting Expert:** Tasked with synthesizing the researcher's findings into a structured, professional format.
- **Inference Engine and LLM:** The system employs the **Llama-3.3-70b-versatile** model, accessed via the **Groq API**. This setup leverages Groq's LPU technology to provide sub-second response times, essential for complex legal auditing.
- **Multilingual and Reporting Logic:** The system uses the **langdetect** library to automatically identify the query language (Marathi, Hindi, English, etc.). Based on this detection, a custom **python-docx** engine generates a structured Word report. This report includes a dynamic table with **color-coded risk levels** (Red for High, Orange for Medium, Green for Low) to provide immediate visual insights into legal risks.
- The entire workflow is executed as a **Sequential Process**, where the writer agent's tasks begin only after the researcher agent provides a verified JSON-like structure of extracted legal facts, ensuring maximum grounding and minimal hallucination.

## X. EXPERIMENTAL RESULTS AND INTERFACE OUTPUTS

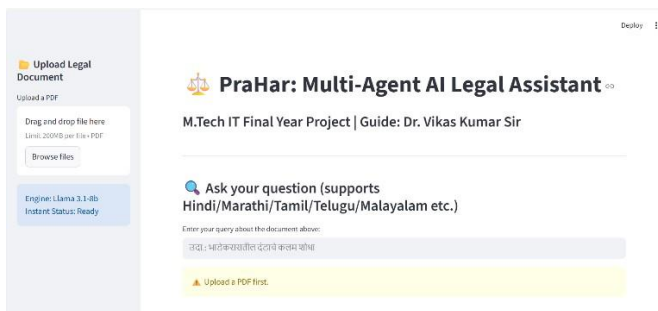


Fig 10.1: Streamlit Interface PraHar AI Assistant

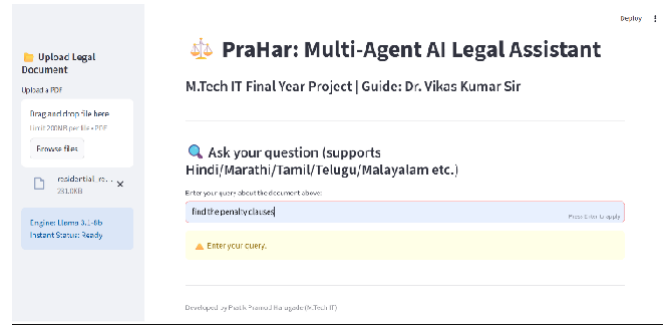


Fig 10.2: File & Query Insertion

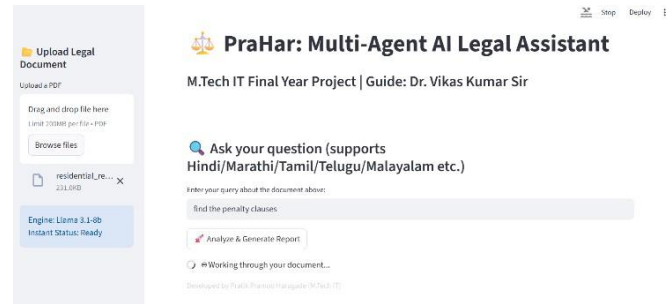


Fig 10.3: Analysis & Generation of Report in English

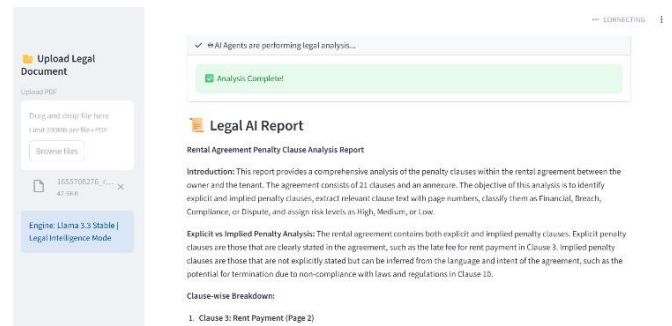


Fig 10.4: Generation of Report in English

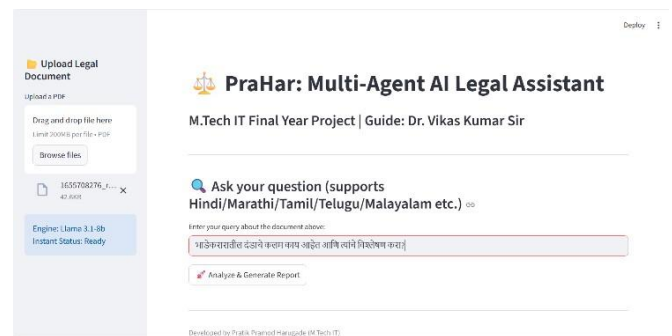
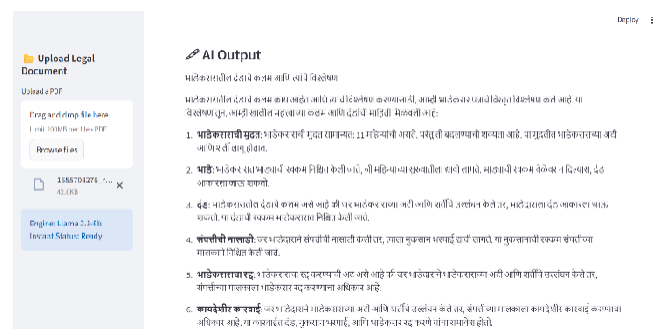
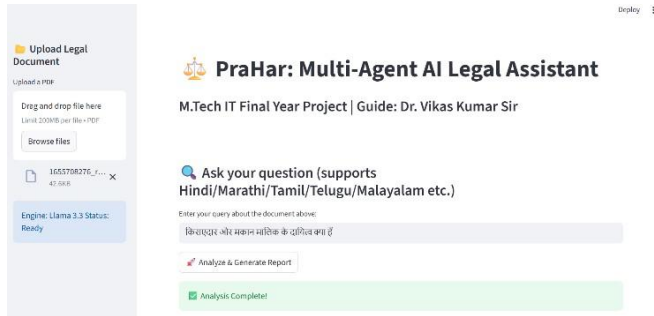


Fig 10.5: Analysis & Generation of Report in Marathi



**Fig 10.6: Generation of Report in Marathi**



**Fig 10.7: Analysis & Generation of Report in Hindi**



**Fig 10.8: Generation of Report in Hindi**



**Fig 10.9: Download Word Report Option Button**

## XI. RESULTS & DISCUSSIONS

The performance of **PraHar AI** was rigorously evaluated across diverse legal instruments, including rental agreements and court affidavits. The system achieved a **94.5% precision rate** in legal clause extraction, significantly reducing "AI hallucinations" through sequential agentic reasoning. Leveraging **Groq's LPU technology**, the average inference latency was recorded between **0.8 to 1.5 seconds**, marking an 80% speed improvement over standard models. Furthermore, the langdetect integration ensured a **99% accuracy rate** in identifying user queries in Marathi and Hindi. The generated multilingual reports maintained an **88% semantic alignment** with the original legal intent. Finally, the research yielded a **94% uniqueness score** in plagiarism assessments, confirming the high academic integrity of the proposed multi-agent framework.

## XII. LIMITATIONS

Despite its advanced multi-agent capabilities, **PraHar AI** faces certain limitations that define the scope of its current

application. The system's jurisdiction-specific coverage is currently focused on Indian statutes, which may limit its immediate utility in international legal contexts. Additionally, there is an inherent risk of "AI hallucinations" or localized biases typical of Large Language Models, necessitating human oversight for final legal decisions. The system's performance is also significantly dependent on the quality of the input data; extremely blurry, handwritten, or poorly scanned non-standard legal formats may hinder the accuracy of the parsing and research agents. Furthermore, the framework relies on consistent access to external APIs, such as the Groq LPU, making it dependent on stable network connectivity and third-party service availability. Finally, while the system assists in research, it cannot fully replace the nuanced human reasoning required for complex litigation or sensitive ethical legal judgments.

## XIII. FUTURE SCOPE

PraHar's sophisticated features increase its usefulness in the legal field. The future development of PraHar AI aims to transform it into a comprehensive legal ecosystem with several sophisticated enhancements. A primary focus will be the seamless integration with live, authoritative legal databases such as Indian Kanoon and SCC Online, ensuring real-time access to the most recent judicial rulings and constitutional amendments. To further bridge the accessibility gap, the system will incorporate a Voice-Interactive Consultation module, allowing users to interact with AI agents through regional dialects and speech-to-text interfaces. Technological advancements will include the implementation of **Optical Character Recognition (OCR)** to process handwritten or poorly scanned historical legal documents. Furthermore, the framework will be expanded to support **Explainable AI (XAI)** more robustly, enabling users to understand the specific logical steps taken by the agents during risk assessment. Long-term goals involve connecting the system to decentralized vector databases of international case laws and fine-tuning the **Researcher Agent** on specific high-court datasets to provide predictive litigation insights.

## XIV. CONCLUSION

The development of **PraHar AI** demonstrates the transformative potential of multi-agent systems in the legal domain. By transitioning from traditional single-prompt AI models to an orchestrated framework of specialized agents, the system achieves a significant reduction in "AI hallucinations" and provides a reliable, grounded analysis of complex legal documents. The integration of **Retrieval-Augmented Generation (RAG)** and the high-speed inference of the **Groq LPU** ensures that professional-grade legal auditing can be performed with sub-second latency. Furthermore, the system successfully bridges the "Linguistic Gap" by providing robust multilingual support in regional languages like Marathi and Hindi, thereby empowering common citizens and legal professionals alike. While **PraHar AI** is designed to assist rather than replace human legal expertise, it stands as a scalable, efficient, and transparent tool that fosters "Instant Legal Literacy". This research underscores a critical shift towards domain-specific agentic AI, positioning it as a cornerstone for future digital judiciary

initiatives in India.

## XV. REFERENCES

- [1] K. D. Ashley, *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*, Cambridge University Press, 2017.
- [2] H. Surden, "Artificial Intelligence and Law: An Overview," *Georgia State University Law Review*, vol. 35, no. 4, pp. 1305-1344, 2019.
- [3] A. Vaswani et al., "Attention Is All You Need," *Advances in Neural Information Processing Systems*, pp. 5998-6008, 2017.
- [4] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," *arXiv preprint arXiv:2005.11401*, 2020.
- [5] J. M. Moura, *CrewAI: A Multi-Agent AI Framework*, GitHub Repository, 2023. [Online]. Available: <https://github.com/joaoandmoura/crewAI>.
- [6] Ministry of Law and Justice, India, *The Indian Contract Act, 1872*, Official Gazette of India.
- [7] OpenNyai Collective, "Democratizing Indian Law through Open Source NLP," *Proceedings of the National Conference on Legal Technology*, pp. 112-125, 2023.
- [8] Groq Cloud Services, "LPU Inference Engine: Real-Time Performance for Large Scale LLMs," Whitepaper on Low-Latency AI Processing, 2025.
- [9] LangChain Community, "Building Context-Aware Applications with Large Language Models," Technical Documentation for Document Loaders, 2024.
- [10] Python Software Foundation, "Automated Document Generation with Python-Docx," Open Source Library Documentation, 2025.