AI LEGAL ASSISTANT

Project Synopsis Report

Submitted in partial fulfillment of the requirement of the degree of

BACHELORS OF TECHNOLOGY
in
CSE with Specialization (AIML)
to
K.R Mangalam University
by

ROHIT RAJ–2301730179, NAMAN JOSHI–2301730152

Abhinav Kumar–2301730191, Aditiya Suman–2301730141

Under the supervision of
PRASHANT PANWAR

Department of Computer Science and Engineering
School of Engineering and Technology
K.R Mangalam University, Gurugram- 122001, India
January 2025

INDEX

ABSTRACT

The AI Legal Assistant is a web-based application designed to assist users in analyzing legal documents. Users can upload legal documents along with additional prompts, and the system will summarize the document, extract relevant laws, and highlight critical aspects. Additionally, the system will display similar past cases along with their judgments. Users will also have the ability to interact with the document through a chat interface.

The solution leverages the Gemma2-2b model, which will be fine-tuned using a legal document summarization dataset and an Indian law QnA dataset. To enhance the model's reasoning capabilities, it will also be trained on a commonsense QnA dataset and integrated with a Neo4j knowledge graph. Retrieval-Augmented Generation (RAG) will be used to retrieve relevant legal provisions and past cases, ensuring the assistant provides well-grounded legal insights.

The web application will be developed using HTML, CSS, and JavaScript for the frontend and Flask for the backend. This project aims to simplify legal research and document analysis, making legal information more accessible and efficient.

Keywords: AI Legal Assistant, NLP, Legal Document Summarization, Gemma2-2B, RAG, Chatbot

1. INTRODUCTION

Legal professionals are often burdened with voluminous case files, contracts, and statutes that require precise interpretation. With over 50 million cases pending in Indian courts, the need for accelerated legal research tools has become critical. Traditional legal research is not only labor-intensive but also prone to missing subtle legal interpretations or case precedents.

The AI Legal Assistant provides a comprehensive solution by automating key tasks—summarization, legal citation extraction, precedent retrieval, and contextual chat. By integrating RAG, LLMs, and knowledge graphs, the system ensures that users receive accurate and context-rich legal insights in real-time.

The importance of such a tool cannot be overstated in today's legal landscape, where timely and accurate access to information is vital. Legal students, practitioners, and even laypersons often struggle with complex jargon and procedural ambiguity. The AI assistant reduces the cognitive load on users by breaking down complex language and offering meaningful summaries. Furthermore, by incorporating frequently updated datasets, the tool keeps pace with legislative amendments and evolving case law, making it a valuable resource for continuous learning and case analysis.

2. MOTIVATION

Manual legal analysis is slow, error-prone, and resource-intensive, especially in time-sensitive cases. Legal documents often contain archaic terminology and complex structures, posing challenges for law students and new practitioners. Existing solutions lack user-friendliness or are too expensive for academic use.

AI offers a powerful opportunity to democratize access to legal knowledge, enhance decision-making, and reduce the time required for legal research. The rise of affordable and open-source AI frameworks has also made it feasible to deploy intelligent solutions at scale, even for small firms or educational institutions.

Moreover, with India's expanding legal ecosystem and digital transformation initiatives, tools that streamline legal document processing can support broader objectives like judicial efficiency and public access to justice. The AI Legal Assistant aligns with these national goals while also encouraging innovation in legal education and practice. It acts as a digital legal aid, especially valuable in rural or underserved areas, where legal expertise may be limited.

3. LITERATURE REVIEW

- Legal Summarization Models: Transformer-based NLP models (e.g., BART, GPT) have shown success in summarizing lengthy legal opinions while preserving key facts and context.
- Case Law Retrieval: Systems like IBM Watson Legal and Westlaw Edge use NLP and semantic search for precedent retrieval.
- Graph-Based Reasoning: Legal knowledge graphs (e.g., Neo4j) have improved the ability of systems to understand relationships between statutes, entities, and events.
- AI Chatbots for Law: Tools such as DoNotPay and LexisBot offer chatbot-based support but often lack case-specific contextual reasoning.

While Western jurisdictions have seen a proliferation of AI-assisted legal research tools, the Indian market remains relatively underexplored. Much of the innovation remains inaccessible to the broader legal community due to cost, licensing, or lack of regional focus. Additionally, many tools operate as black boxes with minimal transparency into their decision-making processes, raising questions about trustworthiness. The integration of explainable AI in the proposed assistant addresses this by providing traceable references and logic behind each insight.

4. GAP ANALYSIS

Most legal AI tools specialize in either summarization or case retrieval—rarely both. Indian legal content is underrepresented in training datasets, reducing relevance. Existing chat-based tools are limited to basic Q&A and do not use knowledge graphs or citation tracking.

Lack of structured, semantic understanding in existing tools affects the quality of insights provided to users. Furthermore, many commercial tools offer limited support for regional languages and do not accommodate the diverse legal practices across India's states. This one-size-fits-all approach fails to address localized needs.

There's also a significant usability gap. Most legal professionals are not trained in data science or NLP, making technical tools difficult to adopt. Our solution prioritizes usability with an intuitive interface, real-time document insights, and actionable legal summaries tailored to both technical and non-technical users. It fills the void by offering an integrated, affordable, and user-friendly platform focused on Indian jurisprudence.

5. PROBLEM STATEMENT

Legal professionals and students need a reliable and intelligent assistant that can:

- Efficiently summarize complex legal documents
- Extract and interpret relevant statutory citations
- Retrieve and display related Indian judgments
- Support interactive document exploration through natural language queries

Current solutions lack a holistic approach combining these capabilities within one platform. Moreover, most tools do not cater to the structural and linguistic nuances of Indian legal documents, limiting their effectiveness in the local context. The absence of explainability, scalability, and domain-specific optimization hinders widespread adoption. This project aims to solve these limitations by offering an AI-based platform trained specifically on Indian legal content, enhanced with semantic reasoning and multilingual support.

6. OBJECTIVES

- Develop an AI-powered **legal document summarization and analysis tool**.
- Implement **RAG-based retrieval** to cite laws and fetch relevant past cases.
- Train a fine-tuned **Gemma2-2b model** for **legal document QnA and summarization**.
- Provide a **chat interface** for interactive document queries.
- Ensure the system is optimized for **Indian legal frameworks**.

7. TOOLS/PLATFORMS USED

Programming Languages & Frameworks:

- Python (Model Training, Backend API)
- JavaScript, HTML, CSS (Frontend Development)
- Flask (Backend Development)

  Gradio, Streamlit

Machine Learning & AI:

- Phi-2.7B (LLM for Legal NLP Tasks)
- RAG (Retrieval-Augmented Generation) (For Law Citations & Case Retrieval)

8. METHODOLOGY

1. **Data Collection & Preprocessing**
    - Curate legal document datasets for summarization fine-tuning.
    - Collect Indian law QnA datasets for improved legal query handling.
    - Structure past case databases for retrieval tasks.
2. **Fine-Tuning the Gemma2-2b Model**
    - Train on legal summarization and Indian law QnA datasets.
    - Fine-tune reasoning capabilities using **commonsense QnA datasets**.
3. **Retrieval-Augmented Generation (RAG) Implementation**
    - Use **two-document RAG** approach:
        - **User-uploaded document** for direct analysis.
        - **Pre-stored legal document** for law citation retrieval.
    - Utilize **prompt engineering** for accurate law citation generation.
4. **Case Law Retrieval System**
    - Implement a search mechanism for retrieving relevant past cases based on the uploaded document.
5. **Web Application Development**

- Build a **responsive frontend** for document upload and chat interaction.
- Develop a **Flask backend** to process documents and handle AI responses.
- Integrate AI-generated results into the web interface dynamically.

9.REFERENCES

1. **Legal Document Summarization Techniques**, IEEE Transactions on Artificial Intelligence, 2023.
2. **Advancements in Retrieval-Augmented Generation**, ACM Computing Surveys, 2024.
3. **Indian Law QnA Dataset**, Open Legal Data Project, 2023.