AI-DRIVEN LEGAL CHATBOT FOR MINING REGULATIONS

Mohammed Azeem A
School of CSE,
Presidency University,
Bangalore, India.
amohdazeem187@gmail.com

Ashfaq Ur Rahman H N
School of CSE,
Presidency University,
Bangalore, India.
ashfaq28rahman@gmail.com

Sagar H N
School of CSE,
Presidency University,
Bangalore, India.
sagarhns817@gmail.com

ABSTRACT: The mining sector in India is governed by complex and evolving regulations that can be challenging for stakeholders to interpret and apply accurately. This research presents an AI-driven legal chatbot specifically developed to address this problem by providing instant, contextual, and accurate answers related to Indian mining laws. The system leverages Natural Language Processing (NLP) and a vector similarity search mechanism to retrieve relevant legal sections from authoritative documents such as the *Mines and Minerals (Development and Regulation) Act, 1957*. Using OpenAI's GPT model as the language engine and FAISS for efficient semantic retrieval, the chatbot enables users to interact through natural language queries and receive concise legal explanations grounded in official legal texts. This solution democratizes access to regulatory knowledge, minimizes legal ambiguity, and enhances decision-making across the mining industry. The system was tested using real regulatory documents and demonstrates promising potential as a 24/7 accessible legal assistant for mining operators, government agencies, and legal professionals.

1. INTRODUCTION

The mining industry in India operates within a complex framework of regulations, primarily governed by statutes such as the Mines and Minerals (Development and Regulation) Act, 1957, along with numerous rules, circulars, and notifications. Navigating these extensive legal texts can be time-consuming and error-prone, especially for stakeholders such as mine operators, legal consultants, and government agencies who require precise and timely information for decision-making and compliance.

To address these challenges, this research proposes an AI-driven legal chatbot tailored to the Indian mining sector. The system enables users to query mining laws core architecture utilizes OpenAI's GPT language model for generating human-like responses, while FAISS (Facebook AI Similarity Search) handles semantic retrieval of relevant legal sections from a preprocessed vector index of legal documents.

To ensure efficient and scalable interaction between the frontend and backend, the system is built using FastAPI, a modern, high-performance web framework for Python. FastAPI enables rapid development of asynchronous APIs and seamless integration with the NLP model, making the chatbot highly responsive and production-ready. The frontend, developed as a web interface, communicates with the FastAPI backend to fetch structured responses enriched by context from the legal corpus.

By combining artificial intelligence with robust backend architecture, the chatbot enhances accessibility to legal information, reduces dependency on manual legal consultations, and empowers users to interpret mining regulations with confidence and ease.

2. EXISTING METHODS

Traditionally, stakeholders in the mining industry rely on manual methods to access and interpret legal information. This typically involves navigating through government websites, downloading lengthy PDFs, and scanning large volumes of text to locate relevant clauses. Even with digital access, the process is tedious and time-consuming due to the lack of intelligent search capabilities and contextual linking. Users must possess prior legal knowledge to identify the applicable sections or Acts, and even minor queries can demand extensive reading or expert consultation.

This manual approach slows down decision-making and increases the risk of misinterpretation, particularly in time-sensitive or compliance-critical scenarios.

DRAWBACKS: Existing methods for accessing mining regulations are slow and inefficient. Manual searches through lengthy PDFs or websites are time-consuming, and traditional search functions lack context. Legal language requires prior knowledge, and information is scattered across various platforms, making it hard to access. Many documents are also not optimized for mobile or searchable, further limiting accessibility.

3. LITERATURE REVIEW

No.	Paper Title	Methodology	Advantages	Limitations
1	Legal Information Retrieval Using NLP Techniques	Rule-based extraction and keyword search	Easy to implement; suitable for structured documents	Lacks semantic understanding; low accuracy on complex queries
2	Leveraging GPT Models for Legal Text Summarization and QA	Transformer-based models (GPT-2, GPT-3) applied to legal corpora	Context-aware responses; scalable for large datasets	Requires significant compute resources and fine-tuning
3	Chatbots for Legal Aid: A Survey and Implementation Review	Survey of legal bots like DoNotPay and Indian Kanoon	Widely adopted; improved legal awareness	Focused on general law; limited customization for niche domains
4	AI-Based Law Assistant for Statutory Compliance in Indian Industries	Use of LLMs with Indian law datasets and retrieval-based QA architecture	Domain-specific retrieval; faster access to relevant sections	Limited datasets; regulatory updates not always incorporated

Fig 1. Literature Review Table

4. PROPOSED METHOD

The proposed system is a domain-specific AI chatbot built to assist stakeholders in the Indian mining industry by answering queries based on mining laws. The architecture combines document parsing, semantic retrieval, and AI-powered response generation. Below are the core components and key features of the implementation:

1. Document Parsing and Chunking: The system begins by processing official mining regulation documents, such as the Mines and Minerals (Development and Regulation) Act, 1957. Using PyPDF2 and custom scripts, the text is extracted from PDF format. The content is then cleaned and segmented into logical units—such as sections and clauses—based on numbered patterns and headings. Each segment is stored as an object in a structured JSON file (mining_chunks.json) containing fields like section number, title, and content. This JSON serves as the foundational knowledge base.

2. Embedding Generation and Vector Indexing: Each text chunk in the JSON file is passed through OpenAI's embedding model to generate high-dimensional vector representations. These vectors are stored and indexed using FAISS (Facebook AI Similarity Search), which supports fast semantic similarity search. This enables the system to efficiently retrieve the most relevant legal sections in response to a user query.

3. User Query Processing and Context Retrieval: When a user submits a natural language question, the system generates an embedding for the query and uses FAISS to search for the top-k most relevant legal chunks. These retrieved chunks are used to build a contextual prompt, which is passed to OpenAI's GPT language model for answer generation.

4. AI Response Generation with GPT: OpenAI's GPT model receives the user query along with the relevant retrieved legal sections as context. It generates a conversational and legally grounded response, tailored to the structure and tone expected in regulatory interpretation.

5. Backend Deployment with FastAPI: The backend logic, including document loading, FAISS search, and OpenAI interaction, is implemented using FastAPI.

This modern Python web framework provides a lightweight and asynchronous API layer that connects the frontend interface to the core NLP engine.

6. Web-based Chat Interface: The system includes a simple HTML frontend (chatbot.html) that allows users to enter queries and view responses. The frontend communicates with the FastAPI backend via HTTP requests, offering a user-friendly experience.

These stages work in synergy to provide an intelligent legal assistant that can retrieve specific clauses, explain regulatory language, and support legal compliance—all through a conversational interface.

5. OBJECTIVES

The primary objectives of this research are:

1. To develop an AI-powered legal chatbot tailored to the Indian mining industry that provides accurate and contextually relevant answers to user queries based on official legal documents.

2. To automate the extraction and preprocessing of legal texts (e.g., the Mines and Minerals (Development and Regulation) Act, 1957) by converting them into structured JSON format for efficient retrieval.

3. To implement a semantic search mechanism using OpenAI's text-embedding-ada-002 model and FAISS to retrieve the most relevant sections of legal content for a given query.

4. To utilize OpenAI's gpt-3.5-turbo model to generate human-readable responses from retrieved legal context, making complex regulatory language accessible to non-experts.

5. To design a lightweight, responsive web interface supported by a FastAPI backend, enabling real-time interaction between users and the chatbot.

6. To ensure that the system can be extended or adapted for other legal domains or acts beyond the mining sector, promoting scalability and reuse.

6. METHODOLOGY

The development of the AI-driven legal chatbot followed a modular and systematic pipeline. The process was divided into several distinct phases, each responsible for a key task in building the end-to-end system. The methodology is described step-by-step below:

1. Document Parsing and Segmentation: Official legal documents such as the Mines and Minerals (Development and Regulation) Act, 1957 are first collected in PDF format. Using PyPDF2 and a custom Python script (prepare_json.py), the content is extracted, cleaned, and segmented into meaningful units (e.g., sections, sub-sections). These segments are stored in a structured JSON format (mining_chunks.json), where each chunk includes metadata like section numbers and full legal text.

2. Text Embedding with OpenAI: Each chunk of legal text from the JSON file is converted into a high-dimensional vector using OpenAI's embedding model text-embedding-ada-002. This is handled in the generate_index.py script. The embeddings capture semantic meaning and make it possible to match user questions to the most contextually relevant legal sections.

3. Vector Indexing with FAISS: The resulting embeddings are stored in a FAISS (Facebook AI Similarity Search) index, which allows for efficient and fast similarity searches. This enables the system to retrieve the most relevant document chunks for any user query based on vector closeness rather than just keyword matching.

4. User Query Processing: When a user submits a query through the chatbot frontend (chatbot.html), it is passed to the backend (main.py via FastAPI). The query is embedded using the same text-embedding-ada-002 model, and the system searches the FAISS index to retrieve the top-k most relevant legal text chunks.

5. Contextual Answer Generation with GPT: The retrieved legal content, along with the original user query, is formatted into a prompt and passed to OpenAI's gpt-3.5-turbo model. This model generates a coherent, legally grounded, and easy-to-understand response based on the input. The generated answer is then returned to the frontend for display.

6. Backend and Frontend Integration: The system is hosted using FastAPI, which handles user requests, model calls, and response delivery. The frontend is built in HTML and JavaScript, offering a minimal and

responsive interface for query submission and response visualization. Communication between frontend and backend is achieved through standard HTTP POST requests.

7. ARCHITECTURE

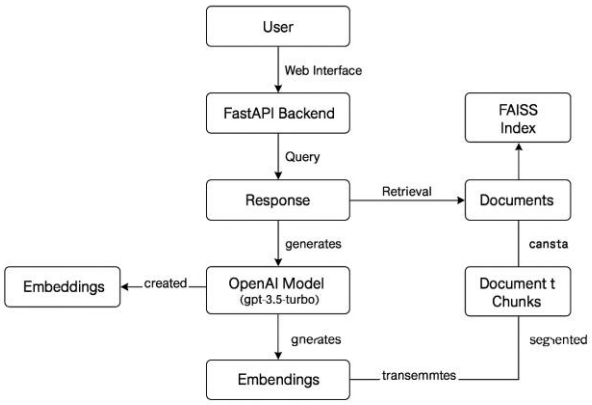

Fig 2. Basic Architecture

8. INTERFACE

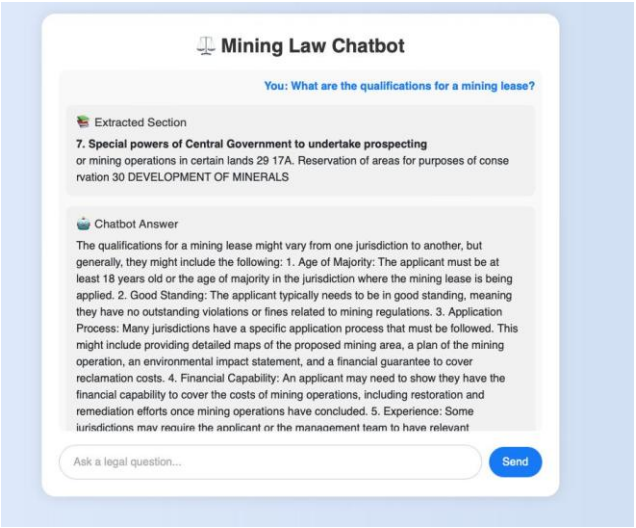

Fig 3. A Web Interface of the ChatBot

9. OUTCOMES

The developed chatbot achieved the following key outcomes:

1. Successfully answered user queries using relevant sections from the Mines and Minerals Act, 1957.

2. Automated the extraction and structuring of legal documents into JSON for efficient retrieval.

3. Implemented accurate semantic search using text-embedding-ada-002 and FAISS.

4. Generated clear and legally sound responses using gpt-3.5-turbo.

5. Provided a responsive and user-friendly interface via FastAPI and a simple web frontend.

6. Offered scalable support for mining law queries with potential for extension to other legal domains.

10. REFERENCES

1. Chiara Valentina Misischia, Flora Poecze, Christine Strauss,Chatbots in customer service: Their relevance and impact on service quality,ISSN18770509,https://doi.org/10.1016/j.procs.2022.03.055.(https://www.sciencedirect.com/science/article/pii/S1877050922004689)

2. A. Behera, K.S. Rawat, S.K. Singh, "Estimation of Mining Subsidence in Talcher Region using Time Series Earth Observation Data", Journal Of The Geological Society Of India 100(8), pg. 1140, (2024); doi:10.17491/jgsi/2024/173962

3. S. Meshram, N. Naik, M. VR, T. More and S. Kharche, "Conversational AI: Chatbots," 2021 International Conference on Intelligent Technologies (CONIT), Hubli, India, 2021, pp. 1-6,doi: 0.1109/CONIT51480.2021.9498508.

4. de Lacerda, Arthur RT, and Carla SR Aguiar. “FLOSS FAQ chatbot project reuse: how to allow nonexperts to develop a chatbot.” Proceedings of the 15th International Symposium on Open Collaboration. 2019.

5. Kohli, Bhaumik, et al. “A Platform for Human-Chatbot Interaction Using Python.” 2018 Second International Conference on Green Computing and Internet of Things (ICGCIoT). IEEE, 2018