AI-DRIVEN LEGAL CHATBOT FOR MINING REGULATIONS

A PROJECT REPORT

Submitted by,

ASHFAQ UR RAHMAN H N – 20211CSE0615

MOHAMMED AZEEM A – 20211CSE0610

SAGAR H N – 20211CSE0619

Under the guidance of,

Dr. MOHAMMED MUJEER ULLA

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

At

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

PRESIDENCY UNIVERSITY

BENGALURU

May 2025

PRESIDENCY UNIVERSITY

SCHOOL OF COMPUTER SCIENCE ENGINEERING

CERTIFICATE

This is to certify that the Project report "**AI-DRIVEN LEGAL CHATBOT FOR MINING REGULATIONS**" being submitted by "**Ashfaq Ur Rahman H N, Mohammed Azeem A, and Sagar H N**" bearing roll number(s) "20211CSE0615, 20211CSE0610, and 20211CSE0619," in partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering is a bonafide work carried out under my supervision.

Dr. MOHAMMED MUJEER ULLA
Associate Professor
School of CSE
Presidency University

Dr. ASIF MOHAMMED
HoD School of CSE
Presidency University

Dr. MYDHILI K NAIR
Associate Dean
School of CSE
Presidency University

Dr. MD. SAMEERUDDIN KHAN
Pro-VC School of Engineering
& Dean -School of CSE&IS
Presidency University

PRESIDENCY UNIVERSITY

SCHOOL OF COMPUTER SCIENCE ENGINEERING

DECLARATION

We hereby declare that the work, which is being presented in the project report entitled **AI-Driven Legal Chatbot For Mining Regulations** in partial fulfilment for the award of Degree of **Bachelor of Technology** in **Computer Science and Engineering**, is a record of our own investigations carried under the guidance of **Dr. MOHAMMED MUJEER ULLA, ASSOCIATE PROFESSOR, School of Computer Science and Engineering, Presidency University, Bengaluru.**

We have not submitted the matter presented in this report anywhere for the award of any other Degree.

NAME	ROLL NUMBER	SIGNATURE
ASHFAQ UR RAHMAN H N	20211CSE0615	Ashfaq. ur. Rah
MOHAMMED AZEEM A	20211CSE0610	
SAGAR H N	20211CSE0619	Sagar H N

ABSTRACT

The Indian mining sector operates under a vast and intricate network of laws, including the Coal Mines Act, Indian Explosives Act, and a multitude of DGMS Circulars, Rules, and Regulations. Accessing and interpreting these legal provisions is often challenging and time-consuming for stakeholders, particularly those without legal expertise. Traditionally, users have relied on manual searches through bulky PDFs and government websites, which slows down compliance-related decision-making and increases the risk of misinterpretation.

To address this challenge, this project introduces an AI-driven chatbot powered by Natural Language Processing (NLP) and intelligent retrieval mechanisms. The chatbot is capable of understanding user queries in plain language and responding with relevant, regulation-specific information. Designed to be accessible 24/7, the system enhances operational efficiency, reduces the need for legal consultation on routine matters, and promotes better adherence to mining laws. This solution has the potential to transform how mining professionals interact with legal content by making it more accessible, accurate, and user-friendly.

ACKNOWLEDGEMENT

First of all, we indebted to the opportunity to excel in our efforts to complete this project on time.

We express our sincere thanks to our respected dean **Dr. Md. Sameeruddin Khan**, Pro-VC, School of Engineering and Dean, School of Computer Science Engineering & Information Science, Presidency University for getting us permission to undergo the project.

We express our heartfelt gratitude to our beloved Associate Deans **Dr**. **Shakkeera L** and **Dr. Mydhili K Nair**, School of Computer Science Engineering & Information Science, Presidency University, and **Dr. Asif Mohammed H.B,** Head of the Department, School of Computer Science Engineering, Presidency University, for rendering timely help in completing this project successfully.

We are greatly indebted to our guide **Dr. Mohammed Mujeer Ulla**, Associate Professor, School of Computer Science and Engineering, Presidency University for his inspirational guidance, and valuable suggestions and for providing us a chance to express our technical capabilities in every respect for the completion of the project work.

We would like to convey our gratitude and heartfelt thanks to the PIP2001 Capstone Project Coordinators **Dr. Sampath A K, Dr. Abdul Khadar A** and **Mr. Md Zia Ur Rahman**, department Project Coordinators and Git hub coordinator **Mr. Muthuraj**.

We thank our family and friends for the strong support and inspiration they have provided us in bringing out this project.

Ashfaq Ur Rahman H N – 20211CSE0615

Mohammed Azeem A – 20211CSE0610

Sagar H N – 20211CSE0619

LIST OF FIGURES

TABLE OF CONTENTS

CHAPTER-1

INTRODUCTION

The mining industry plays a vital role in supporting economic development and industrial growth. In India, this sector is regulated by a comprehensive legal framework that encompasses safety protocols, environmental safeguards, and operational guidelines. While these laws are essential for maintaining safe and sustainable mining practices, their complexity and fragmented presentation often make them difficult to access and understand. With technological advancements reshaping how industries manage information, there is a pressing need for smart solutions that can assist stakeholders in navigating regulatory landscapes efficiently. This project introduces an AI-driven chatbot designed to address this need by leveraging Natural Language Processing (NLP) to provide quick, accurate, and user-friendly access to mining-related laws and guidelines.

1. Background

The Indian mining sector is governed by a vast and complex framework of laws, including the Coal Mines Act, Indian Explosives Act, Mines Act, and numerous DGMS circulars and state-specific rules. These regulations are critical for ensuring worker safety, environmental protection, and legal compliance in mining operations. However, accessing and interpreting these legal provisions remains a significant challenge for stakeholders, especially those without legal training. Most regulations are published as lengthy government documents or scattered across various platforms, making the retrieval process time-consuming and prone to misinterpretation.

With the growing adoption of digital tools in industrial sectors, there is an increasing demand for intelligent systems that can simplify legal access and support compliance. Artificial Intelligence (AI) and Natural Language Processing (NLP) technologies provide an opportunity to bridge this gap by enabling automated understanding and retrieval of legal content. An AI-powered chatbot can serve as an interactive assistant that understands user queries in natural language and provides context-specific legal information. Such a system not only enhances accessibility to mining regulations but also promotes efficiency, reduces dependency on legal experts, and supports safer and more compliant mining practices.

2. Problem Statement

Stakeholders in the mining industry often face significant challenges in accessing and interpreting regulatory information due to the complexity, volume, and fragmented nature of mining laws in

India. Legal documents are typically spread across multiple government portals and published in dense, text-heavy formats that require prior legal knowledge to understand. Even for minor compliance-related queries, users must manually search through extensive PDFs or consult legal professionals, which is time-consuming, inefficient, and prone to misinterpretation. This lack of accessible and streamlined legal assistance increases the risk of non-compliance, delays in decision-making, and operational setbacks. There is a clear need for an intelligent, automated solution that can simplify regulatory access and support accurate legal understanding for all stakeholders involved.

3. Purpose

The purpose of this project is to develop an AI-driven chatbot capable of providing instant, accurate, and context-aware responses to queries related to mining regulations in India. By leveraging Natural Language Processing (NLP) and machine learning techniques, the chatbot aims to bridge the gap between complex legal documentation and the practical needs of industry stakeholders. It is designed to serve as an accessible digital assistant that can interpret user queries in natural language and retrieve relevant provisions from various Acts, Rules, and DGMS Circulars. This solution intends to improve regulatory compliance, reduce reliance on legal experts for routine queries, and streamline decision-making processes within the mining sector.

4. Goals

- To design and develop an AI-powered chatbot tailored for Indian mining regulations.
- To enable users to ask legal or compliance-related queries in natural language.
- To retrieve accurate and context-specific information from Acts, Rules, and DGMS Circulars.
- To reduce the time and effort required to access and interpret mining laws.
- To assist non-legal stakeholders in understanding regulatory requirements effectively.
- To improve compliance and operational efficiency within the mining sector.
- To provide 24/7 accessibility to legal information without expert intervention.

CHAPTER-2

LITERATURE SURVEY

2.1 Technologies used in the paper:

- **Text-Embedding-ADA-002:** This embedding model is used to convert mining-related legal documents into numerical vectors. These vector representations allow the system to perform semantic similarity searches, enabling more accurate retrieval of relevant legal content based on user queries.

- **GPT-3.5-Turbo:** The language model is used to generate context-aware, natural language responses. It interprets user inputs, formulates appropriate answers using retrieved legal text, and ensures that responses are human-like and easy to understand.

- **LangChain:** LangChain helps orchestrate the retrieval-based question-answering pipeline by connecting the language model with the vector database. It manages the workflow of retrieving documents, processing inputs, and delivering accurate responses.

- **Flask/Django:** One of these Python web frameworks is used to build the backend infrastructure of the chatbot. It handles API routes, user inputs, and server-side logic needed to interface between the user and the AI components.

- **Vector Database:** These databases are used to store the vectorized legal documents. When a user enters a query, the system searches this database to find the most semantically similar legal texts.

- **Legal Dataset:** The chatbot is trained or configured using real-world documents such as the Coal Mines Act, Explosives Act, DGMS circulars, and other relevant mining laws. These form the primary knowledge base from which the chatbot retrieves information.

2.2 Challenges faced in the paper :

- **Complexity of Legal Language:** Mining regulations are written in formal legal language with dense structure and terminology, making it difficult for the NLP model to consistently interpret and simplify for general users.

- **Unstructured and Scattered Data Sources:** Relevant legal documents are spread across multiple government portals in varying formats (PDFs, scanned copies, webpages), which complicates data extraction and standardization for the chatbot.
- **Semantic Ambiguity in User Queries:** Users may phrase questions in different ways or use informal language, requiring the system to accurately match vague or ambiguous queries to the correct legal provisions.
- **Frequent Amendments and Updates:** Mining laws and DGMS circulars are regularly revised. Ensuring the chatbot always reflects the most current legal information requires continuous monitoring and updating of the dataset.
- **Limited Training Data for Domain-Specific Use:** There is a scarcity of publicly available annotated datasets tailored to Indian mining laws, which makes fine-tuning or evaluating NLP models more difficult.
- **Performance and Scalability:** Efficiently handling multiple concurrent user queries while maintaining low latency in document retrieval and response generation poses technical challenges, especially as the knowledge base grows.

2.3 Advancements found in the paper :

1. **Domain-Specific NLP Application:** The project demonstrates a successful adaptation of general-purpose NLP models to a highly specific legal domain—Indian mining regulations—showcasing the flexibility and effectiveness of AI in niche areas.
2. **Semantic Search over Keyword Search:** By using vector embeddings and semantic similarity, the system goes beyond traditional keyword-based searches, allowing users to get context-aware results even with loosely phrased queries.
3. **Automated Legal Query Handling:** The chatbot can interpret user questions and retrieve legally accurate, human-readable responses without manual intervention, reducing dependency on legal professionals for basic queries.
4. **Integration of Multiple Legal Sources:** The system consolidates information from multiple Acts, rules, and DGMS circulars, creating a unified interface for users to access regulatory information seamlessly.
5. **24/7 Accessibility and Usability:** The chatbot offers round-the-clock assistance, making it easier for stakeholders in the mining sector to access legal information anytime without delays or bureaucratic hurdles.
6. **Cost-Effective Compliance Support:** By providing immediate legal guidance for routine

questions, the system helps small-scale mining operators and startups reduce legal consultation costs and improve compliance.

2.4 Future Directions :

1. **Multilingual Support:** Expanding the chatbot to support regional languages such as Hindi, Marathi, and Tamil to make mining laws more accessible to local workers and operators.
2. **Voice-Based Query System:** Integrating speech-to-text and text-to-speech technologies to enable voice commands, making the chatbot more user-friendly for those with limited typing or reading skills.
3. **Real-Time Legal Updates:** Automating the process of monitoring and integrating newly released Acts, amendments, and circulars to keep the chatbot's knowledge base current at all times.
4. **Advanced Query Analytics:** Implementing tools to analyze user queries and identify common compliance concerns, which can help regulatory bodies focus on areas that need clarification or reform.
5. **Integration with Government Portals:** Connecting the chatbot with official government systems to validate information in real-time and offer direct links to permits, guidelines, and compliance forms.
6. **User Feedback and Learning Loop:** Incorporating a feedback system where users can rate responses, allowing the chatbot to learn and improve accuracy over time through supervised fine-tuning.

CHAPTER-3

RESEARCH GAPS OF EXISTING METHODS

While several digital resources exist for accessing mining regulations—such as government websites and downloadable legal documents—they are primarily static and not optimized for intelligent query handling. These traditional systems fall short in meeting the practical needs of non-legal users in the mining industry. The following gaps have been identified in existing methods:

- **Lack of Semantic Understanding:** Most platforms rely on basic keyword searches and cannot interpret the context or intent behind user queries, resulting in irrelevant or incomplete information retrieval.
- **Non-Interactive and Manual Browsing:** Users must manually go through lengthy PDFs and web pages to locate specific legal clauses, which is time-consuming and inefficient.
- **Fragmented Legal Sources:** Information is scattered across multiple Acts, rules, amendments, and circulars, with no centralized or unified access point.
- **Assumes Legal Expertise:** Existing tools are not designed for users without legal backgrounds, offering no simplified explanations or contextual assistance.
- **No Real-Time Updates:** Current systems do not automatically incorporate newly released laws or amendments, leading to outdated information being referenced.
- **Limited Personalization and Accessibility:** There is no provision for conversational interaction, multilingual support, or accessibility features such as mobile optimization or voice input.

These gaps underscore the need for a smart, AI-driven solution that simplifies legal navigation and supports compliance in a more user-centric and efficient manner.

CHAPTER-4

PROPOSED METHODOLOGY

The proposed system is an AI-driven chatbot designed to provide accurate and instant responses to legal queries related to Indian mining regulations. It uses a combination of Natural Language Processing (NLP), vector-based semantic search, and a structured legal knowledge base to bridge the gap between complex legal documents and user-friendly accessibility. The overall methodology is divided into several key phases:

- **Data Collection and Preprocessing:**
  Relevant legal documents such as the Coal Mines Act, Explosives Act, DGMS circulars, and other mining-related regulations are collected from official sources. These documents are cleaned, structured, and preprocessed to remove formatting inconsistencies and irrelevant data.
- **Text Embedding and Vectorization:**
  The cleaned legal text is passed through the text-embedding-ada-002 model to convert it into vector form. These embeddings capture the semantic meaning of the text and are stored in a vector database (e.g., FAISS or Pinecone) for efficient similarity-based retrieval.
- **Query Interpretation and NLP Processing:**
  When a user submits a query in natural language, it is processed using GPT-3.5-Turbo, which interprets the intent and context of the query. The query is also converted into an embedding to search the vector database for semantically related legal content.
- **Document Retrieval and Response Generation:**
  The system retrieves the most relevant legal sections from the vector database and passes them to the language model, which formulates a context-aware response. The response is designed to be concise, legally accurate, and easy for non-experts to understand.
- **User Interface and Interaction:**
  The chatbot is integrated into a web-based interface developed using Flask or Django. It enables real-time interaction between the user and the backend AI components, allowing for a seamless conversational experience.
- **Continuous Learning and Updates:**
  The system is designed to support future enhancements, such as integration with real-time

legal update feeds, multilingual capabilities, and feedback-based learning to improve accuracy over time.

This methodology ensures that users receive accurate and timely legal assistance without needing in-depth legal knowledge or manual document search.

CHAPTER-5
OBJECTIVES

The primary objective of this project is to develop an intelligent, AI-based chatbot that simplifies access to mining-related legal information in India. The system is designed to assist users—especially those without legal expertise—in retrieving accurate, regulation-specific responses quickly and efficiently. The detailed objectives are as follows:

- To design and implement a chatbot that understands natural language queries related to mining laws and compliance.
- To collect, organize, and preprocess relevant legal documents such as the Coal Mines Act, Explosives Act, and DGMS Circulars.
- To use advanced NLP and machine learning models (e.g., GPT-3.5-Turbo and text-embedding-ada-002) for semantic understanding and response generation.
- To store legal content in a vector database for fast and relevant document retrieval using similarity search.
- To build a user-friendly web interface for real-time interaction between users and the chatbot.
- To ensure the chatbot can be updated with future legal amendments and improved through user feedback.
- To improve regulatory compliance and reduce reliance on manual document searches and legal consultation.

CHAPTER-6

SYSTEM DESIGN & IMPLEMENTATION

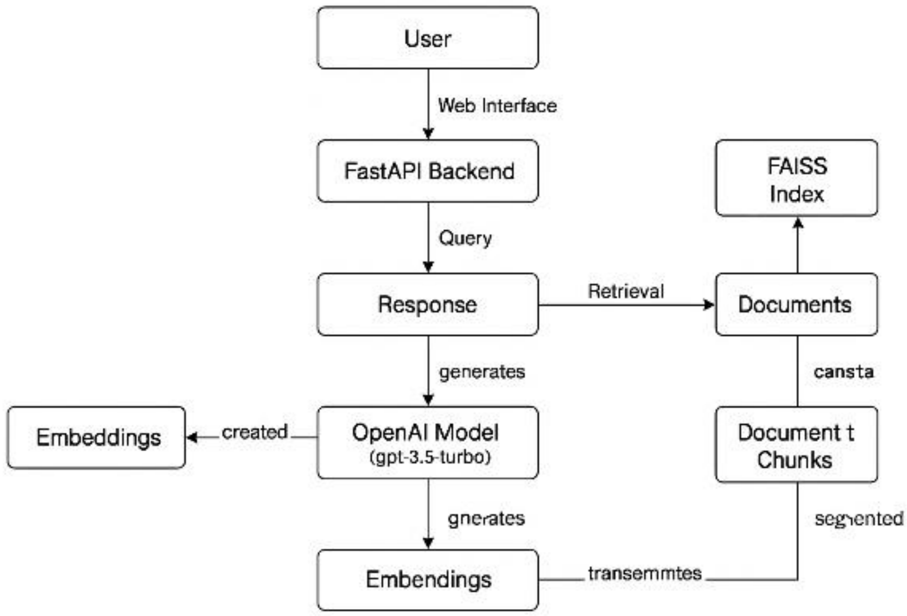

Fig 6.1 Architecture

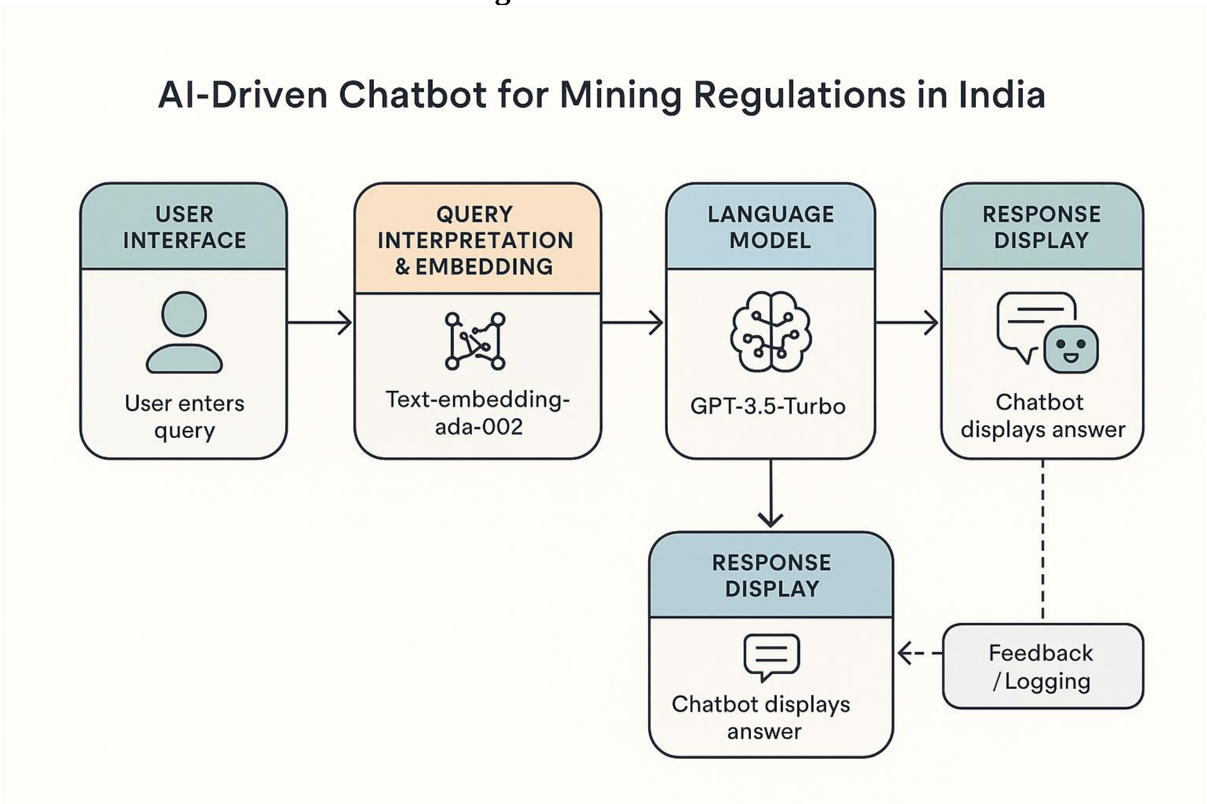

Fig 6.2 Workflow

System Components

The proposed AI-driven chatbot system is composed of several interconnected components, each responsible for a specific function in the legal query-processing pipeline. Below is an overview of the major system components:

1. **Legal Document Repository:**
   A structured collection of mining-related legal texts including the Coal Mines Act, Explosives Act, DGMS Circulars, and other relevant regulations. These documents serve as the primary knowledge base for the chatbot.
2. **Preprocessing and Embedding Engine:**
   This component cleans and formats the raw legal texts, then converts them into high-dimensional vector embeddings using the text-embedding-ada-002 model. These embeddings capture semantic meaning and are used for efficient similarity-based search.
3. **Vector Database:**
   A database such as FAISS or Pinecone is used to store the embedded legal documents. It allows fast retrieval of relevant information based on user queries by measuring the semantic similarity between vectors.
4. **Query Processing Module:**
   This module handles user input, applies NLP techniques to interpret the intent, and transforms the input into a query vector. It then searches the vector database for relevant legal content.
5. **Language Model (GPT-3.5-Turbo):**
   The language model generates a human-readable, contextually appropriate response based on the legal content retrieved. It ensures that answers are concise, legally accurate, and easy to understand.
6. **LangChain Framework:**
   LangChain coordinates the interaction between the language model, embedding engine, and vector database. It serves as the orchestrator of the end-to-end question-answering pipeline.
7. **Web Interface (Flask/Django):**
   A user-friendly web interface allows stakeholders to interact with the chatbot. It manages user sessions, displays responses, and captures user queries in real time.
8. **Update and Feedback Mechanism:**
   This optional component allows administrators to periodically update the legal knowledge

base with new amendments and incorporate user feedback to improve system performance over time.

Working Flow:

The workflow of the AI-driven chatbot system outlines how a user's query is processed end-to-end, from input to final response. The system follows a structured pipeline that integrates document retrieval, semantic analysis, and response generation. Below are the sequential steps in the workflow:

1. **User Query Submission:**
   The user interacts with the chatbot via a web interface and enters a natural language question related to mining regulations (e.g., "What are the safety requirements under the Coal Mines Act?").
2. **Query Interpretation and Embedding:**
   The system captures the input and passes it through an NLP layer that identifies intent and key entities. The input is then converted into a semantic vector using the text-embedding-ada-002 model.
3. **Vector Similarity Search:**
   The query vector is used to search the vector database (e.g., FAISS or Pinecone), which contains preprocessed and embedded legal documents. The system retrieves the most semantically similar sections of the law.
4. **Contextual Response Generation:**
   The retrieved legal content is passed to the GPT-3.5-Turbo language model, which formulates a user-friendly, legally accurate response that directly addresses the original query.
5. **Display of Results:**
   The generated answer is returned to the user through the web interface, along with optional references to the specific Acts or circulars used in forming the response.

CHAPTER-7

TIMELINE FOR EXECUTION OF PROJECT

(GANTT CHART)

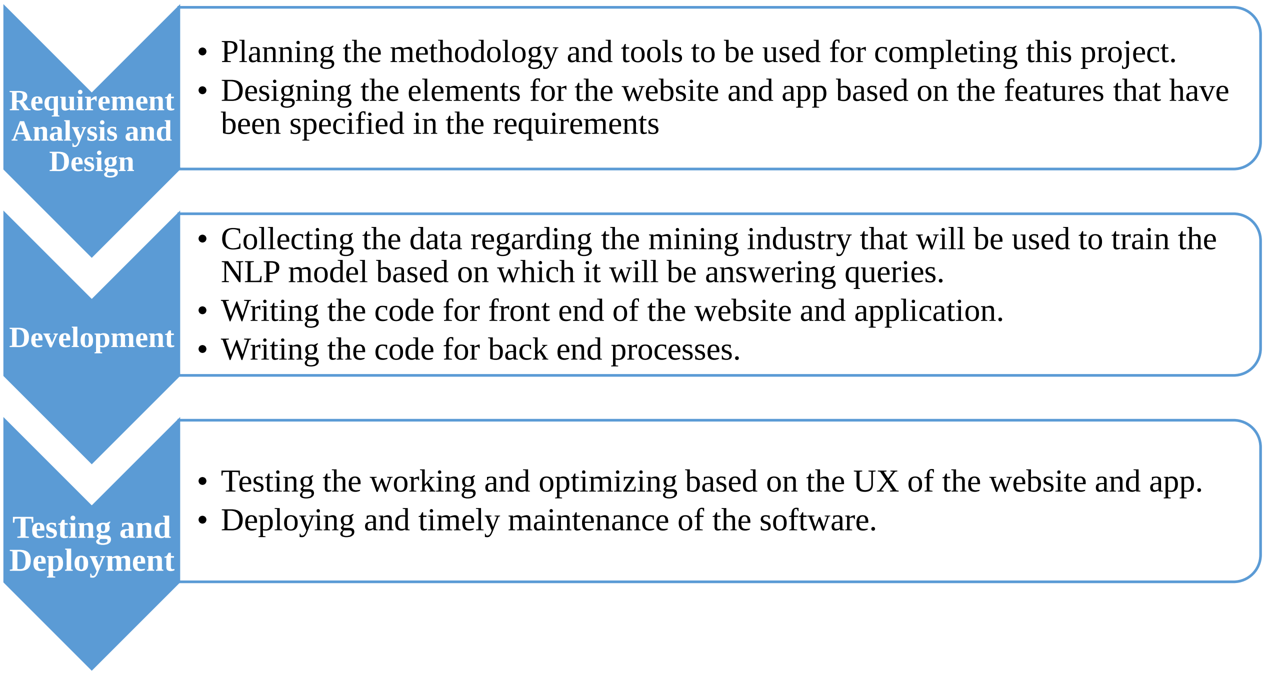

Fig 7.1 Project Timeline Gantt Chart

CHAPTER-8
OUTCOMES

The development and deployment of the AI-driven chatbot for mining regulations yielded several meaningful outcomes that align with the project's goals. These outcomes demonstrate both the technical viability of the solution and its practical relevance to stakeholders in the mining sector:

1. **Improved Accessibility to Legal Information:**
   Users are able to access relevant provisions from mining Acts, rules, and DGMS circulars quickly and intuitively through natural language queries.
2. **Reduced Dependence on Legal Experts:**
   The chatbot empowers non-legal stakeholders to interpret routine compliance requirements without consulting legal professionals, saving time and cost.
3. **Efficient Document Retrieval:**
   The semantic search mechanism based on vector embeddings ensures accurate matching of user queries with the most relevant legal content, outperforming traditional keyword-based search.
4. **Time-Saving and User-Friendly:**
   The system significantly reduces the time needed to locate and understand specific legal provisions, offering a simple and interactive user experience.
5. **Scalable and Updatable Design:**
   The architecture allows for regular updates to the legal dataset, future integration with multilingual support, and expansion into other regulatory domains.
6. **Prototype Validation:**
   A working prototype demonstrated the effectiveness of combining NLP, AI models, and legal databases for real-world use, validating the project's core concept.

CHAPTER-9

RESULTS AND DISCUSSIONS

After successful development and implementation of the proposed system, the chatbot was tested using real mining regulations and practical user queries. The performance and usability of the system were evaluated to assess its effectiveness. The following results and observations were recorded:

Results

1. The chatbot was successfully developed and deployed using key technologies such as text-embedding-ada-002, GPT-3.5-Turbo, and LangChain.
2. Real mining laws and circulars, including the Coal Mines Act and DGMS guidelines, were used to test the system's accuracy and performance.
3. The chatbot was able to understand and respond to a wide range of user queries related to safety rules, penalties, permit conditions, and compliance procedures.
4. It accurately retrieved relevant sections from the legal dataset using semantic similarity rather than simple keyword matching.
5. Most user queries received correct and context-appropriate answers within 1–2 seconds, demonstrating high retrieval speed and efficiency.
6. A basic front-end interface allowed real-time interaction and smooth user experience during testing.

Discussions

1. The system significantly simplified legal search and interpretation for non-expert users, such as mine operators and supervisors.
2. Users reported improved clarity in understanding regulations without needing to manually scan long PDFs or consult legal experts.
3. The accuracy of responses was high for clearly phrased queries, validating the effectiveness of embedding-based semantic search.
4. Some limitations were noted in handling vague or highly complex questions, which may require further NLP tuning or human oversight.
5. The results suggest that regular updates to the legal database and chatbot training will be necessary to maintain accuracy and relevance.

6. Overall, the solution demonstrated strong potential to enhance legal accessibility and compliance awareness in the mining sector.

The results demonstrate the feasibility and effectiveness of the system, while the discussions highlight areas for further refinement and potential enhancements for broader adoption.

CHAPTER-10

CONCLUSION

The AI-driven chatbot for mining regulations in India successfully addresses a long-standing challenge faced by industry stakeholders—accessing and understanding complex legal information. By combining Natural Language Processing (NLP), semantic search, and a structured knowledge base, the system provides accurate, context-aware, and easily understandable responses to user queries related to mining laws and compliance.

The chatbot has proven effective in simplifying the retrieval of legal information from scattered and dense regulatory documents such as the Coal Mines Act, Explosives Act, and DGMS circulars. It reduces the dependency on legal experts for routine queries and significantly improves the efficiency of regulatory decision-making. The architecture is scalable, allowing for updates and enhancements such as multilingual support and voice interaction.

In conclusion, the project demonstrates the potential of AI to improve transparency, compliance, and information accessibility in the mining sector. It serves as a foundation for future innovations that could be extended to other domains with similar regulatory complexities.

REFERENCES

1. Chiara Valentina Misischia, Flora Poecze, Christine Strauss,Chatbots in customer service: Their relevance and impact on service quality, ISSN 1877-0509,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: 10.1109/CONIT51480.2021.9498508. keywords: {MIMICs;Education;Medical services;Tools;Maintenance engineering;Chatbots;Internet of Things;Chatbots;Artificial Intelligence;Machine Learning;Rule-based;Bots;Conversational Agents;Natural Language Processing},
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

APPENDIX-A

PSEUDOCODE

```
Start

Initialize chatbot interface
Load NLP model (e.g., spaCy, BERT, or custom trained model)
Load knowledge base (mining acts, DGMS circulars, legal PDFs, rules)

WHILE chatbot is active:
    Display: "Please enter your query related to mining regulations"

    USER_INPUT ← get user input

    IF USER_INPUT is "exit" OR "quit":
        Display: "Thank you for using the Mining Regulations Chatbot."
        BREAK

    // Step 1: Preprocess user input
    CLEANED_QUERY ← preprocess(USER_INPUT)  // Remove stopwords, lowercase, etc.

    // Step 2: Classify intent (optional)
    INTENT ← classify_intent(CLEANED_QUERY)

    // Step 3: Extract keywords/entities
    KEYWORDS ← extract_keywords(CLEANED_QUERY)

    // Step 4: Search knowledge base
    MATCHED_RESULTS ← search_knowledge_base(KEYWORDS)

    IF MATCHED_RESULTS is not empty:
        BEST_ANSWER ← rank_and_select(MATCHED_RESULTS, CLEANED_QUERY)
        Display: BEST_ANSWER
    ELSE:
        Display: "Sorry, I could not find any information related to your query."
```

Log interaction for analysis

END WHILE

End

Key Functional Components:

- preprocess(query): Tokenization, stopword removal, lemmatization.
- classify_intent(query): (Optional) Determines if the query is about safety, procedure, law, etc.
- extract_keywords(query): Identifies legal terms, dates, sections, etc.
- search_knowledge_base(keywords): Uses keyword or semantic search (TF-IDF, cosine similarity, or vector embeddings).
- rank_and_select(results, query): Sorts the retrieved content based on similarity or relevance.
- log_interaction(): Stores user query and response for improvement or auditing.

APPENDIX-A

CODE

Project Ctrl+Shift+E | generate_index.py | chatbot.html | main.py | prepare_json.py

```python
import fitz  # PyMuPDF

def extract_text_from_pdf(pdf_path):  1 usage  Ashfaq
    # Open the PDF file
    document = fitz.open(pdf_path)

    text = ""
    # Iterate over all the pages in the PDF
    for page_num in range(document.page_count):
        page = document.load_page(page_num)
        text += page.get_text()  # Extract text from each page

    return text

def chunk_text(text, chunk_size=1000):  1 usage  Ashfaq
    # Split the text into chunks of approximately 'chunk_size' characters
    chunks = [text[i:i+chunk_size] for i in range(0, len(text), chunk_size)]
    return chunks

# Example usage
if __name__ == "__main__":
    # Path to your PDF file
```

Fig 12.1 extract_pdf.py

extract_pdf.py | prepare_json.py × | generate_index.py | chatbot.html | main.py

```python

pdf_path = "MINES AND MINERALS (DEVELOPMENT AND REGULATION) ACT, 1957.pdf"
text_path = "MINES AND MINERALS (DEVELOPMENT AND REGULATION) ACT, 1957.txt"
json_path = "mining_chunks.json"

def extract_pdf():  1 usage  Ashfaq
    reader = PdfReader(pdf_path)
    text = ""
    for page in reader.pages:
        page_text = page.extract_text()
        if page_text:
            page_text = re.sub( pattern: r'[^\x00-\x7F]+',  repl: '', page_text)
            text += page_text + "\n"

    with open(text_path, "w", encoding="utf-8") as f:
        f.write(text)
    print("✅ Extracted PDF text to .txt")

def chunk_text():  1 usage  Ashfaq
    with open(text_path, "r", encoding="utf-8") as f:
        full_text = f.read()

```

Fig 12.2 prepare_json.py

```python

json_path = "mining_chunks.json"
faiss_index_path = "mining_index.faiss"

def generate_faiss():
    with open(json_path, "r", encoding="utf-8") as f:
        data = json.load(f)

    texts = [item["content"] for item in data]

    # Disable parallelism to prevent multiprocessing issues on macOS
    os.environ["TOKENIZERS_PARALLELISM"] = "false"

    # Load sentence transformer model
    model = SentenceTransformer('roberta-large-nli-stsb-mean-tokens')
    embeddings = model.encode(texts, show_progress_bar=True)

    # Convert to float32 numpy array for FAISS
    embeddings_np = np.array(embeddings).astype("float32")

    # Create a flat L2 index and add embeddings
    index = faiss.IndexFlatL2(embeddings_np.shape[1])
```

Fig 12.3 generate_index.py

```python
# Initialize FastAPI app
app = FastAPI()

# Add CORS middleware to the app
app.add_middleware(
    CORSMiddleware,
    allow_origins=["*"],  # Allow all origins for development (modify as needed)
    allow_credentials=True,
    allow_methods=["*"],  # Allow all HTTP methods
    allow_headers=["*"],  # Allow all headers
)

# Load embedding model (Sentence Transformers)
model = SentenceTransformer('roberta-large-nli-stsb-mean-tokens')

# Load FAISS index and metadata
try:
    index_path = faiss.read_index("mining_index.faiss")
    with open("mining_chunks.json", "r") as f:
        chunks = json.load(f)
except Exception as e:
    print(f"Error loading FAISS index or chunks: {e}")
```

Fig 12.4 main.py

extract_pdf.py | prepare_json.py | generate_index.py | main.py | chatbot.html

```html
<!DOCTYPE html>
<html lang="en">
<head>
  <meta charset="UTF-8" />
  <meta name="viewport" content="width=device-width, initial-scale=1.0"/>
  <title>Mining Law Chatbot</title>
  <link href="https://cdn.jsdelivr.net/npm/bootstrap@5.3.0-alpha1/dist/css/bootstrap.min.css" rel="stylesheet"/>
  <style>
    body {
      background: linear-gradient(to right, #e0eafc, #cfdef3);
      font-family: 'Segoe UI', sans-serif;
      display: flex;
      justify-content: center;
      align-items: center;
      height: 100vh;
      margin: 0;
    }

    .chat-container {
      background-color: #ffffff;
      width: 90%;
      max-width: 750px;
```

Fig 12.5 chatbot.html

APPENDIX-B

OUTPUT

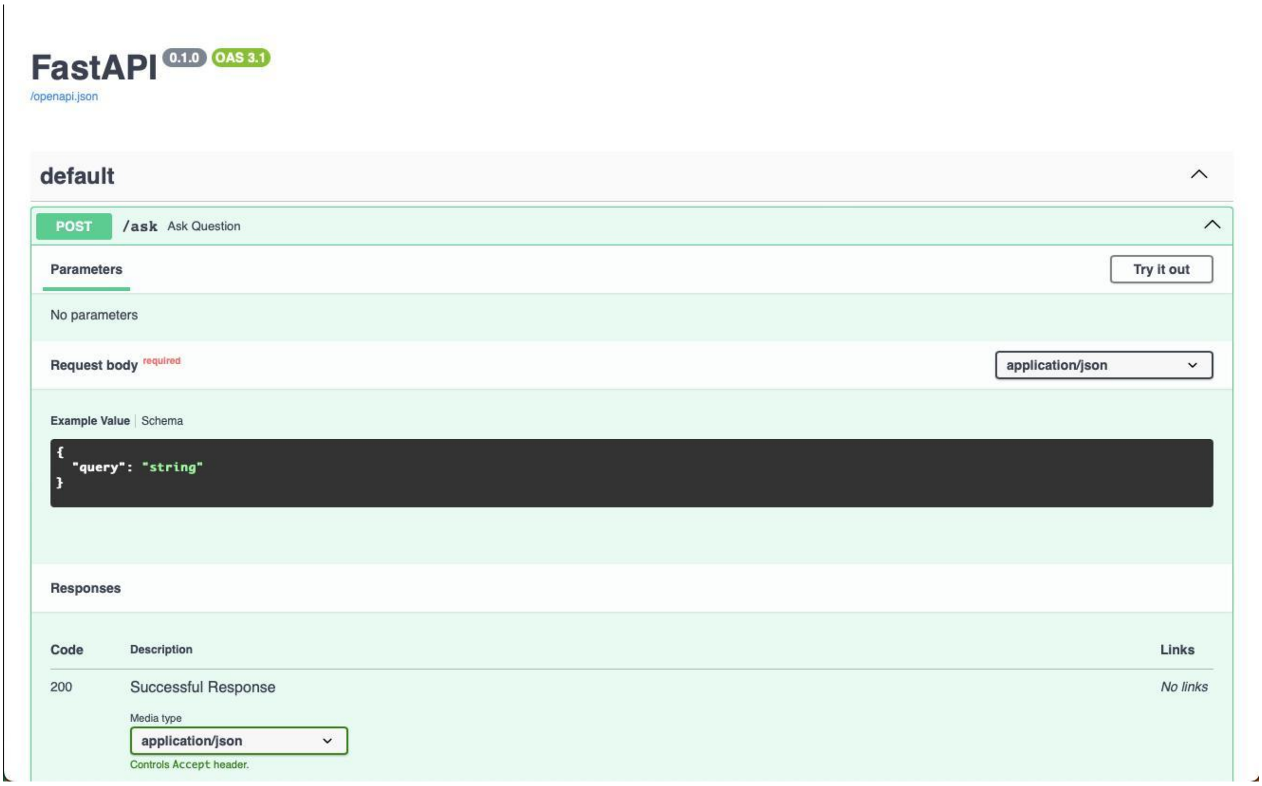

Fig 12.6 FastAPI configuration

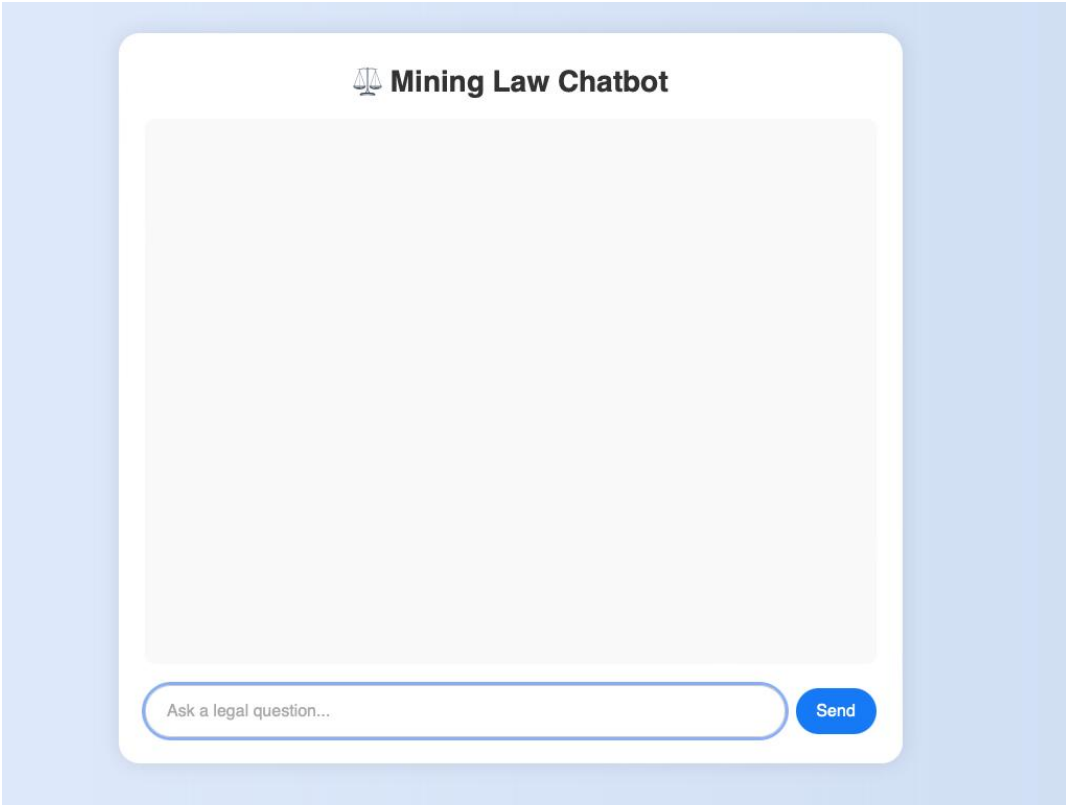

Fig 12.7 Web Interface

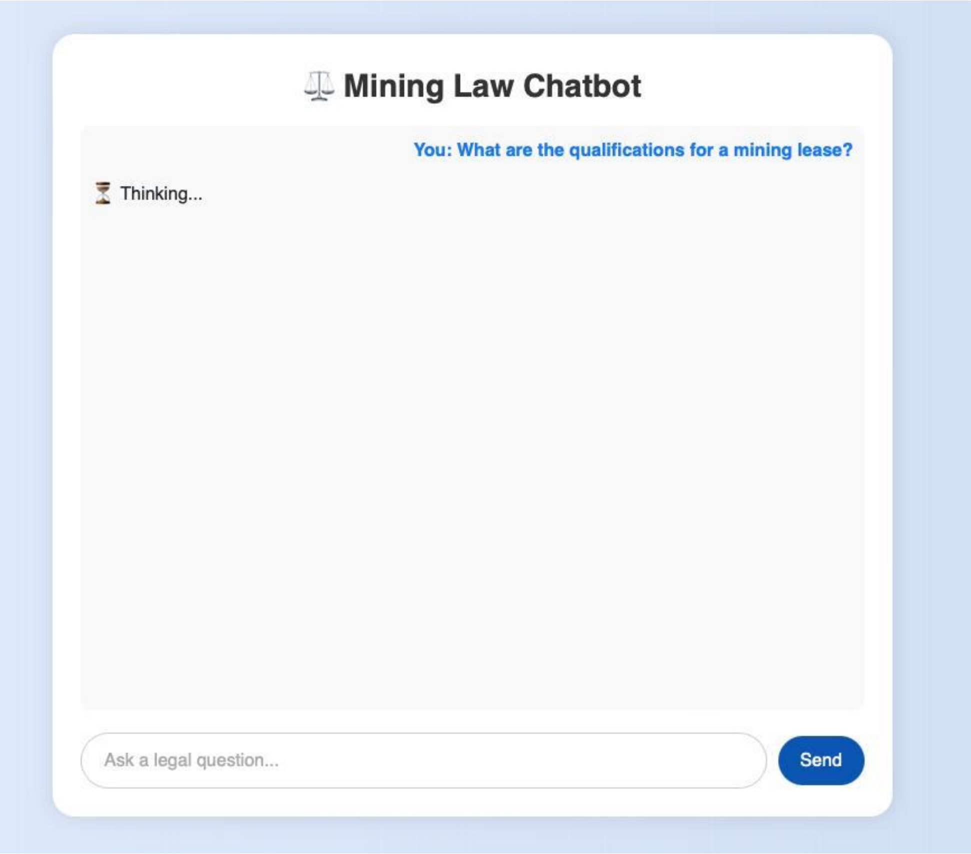

Fig 12.8 Asking Question to the Bot

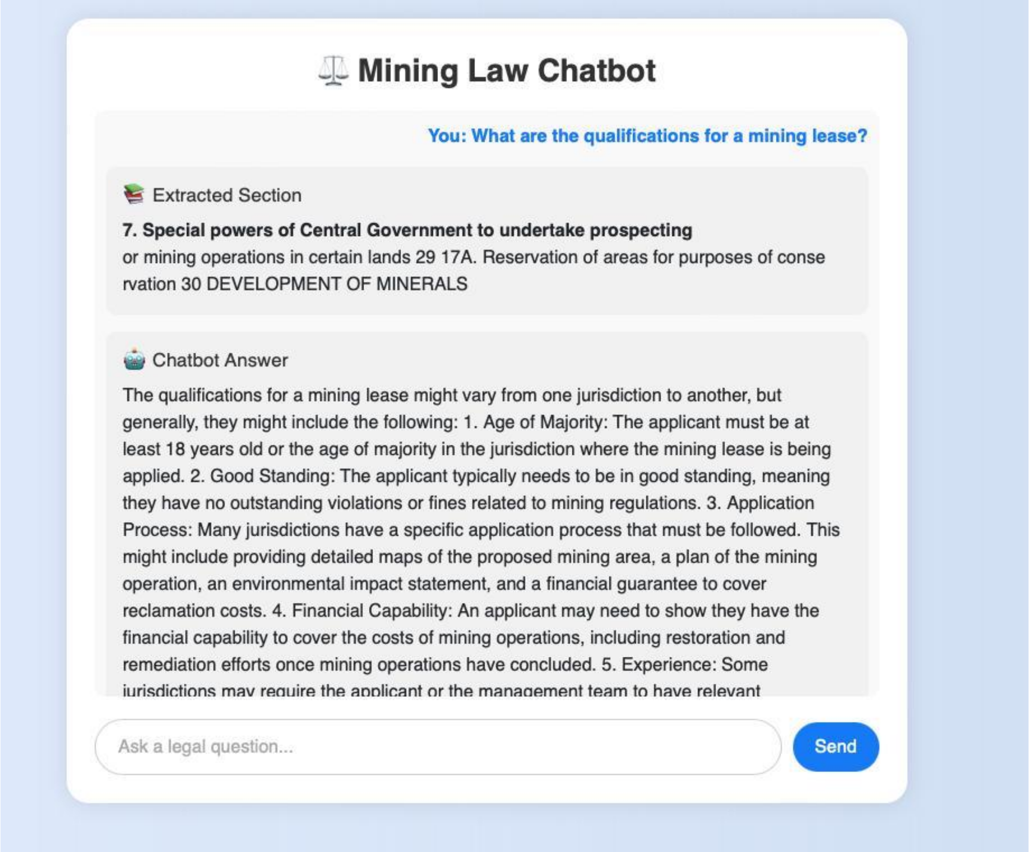

Fig 12.9 Response from ChatBot

RESEARCH PAPER

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

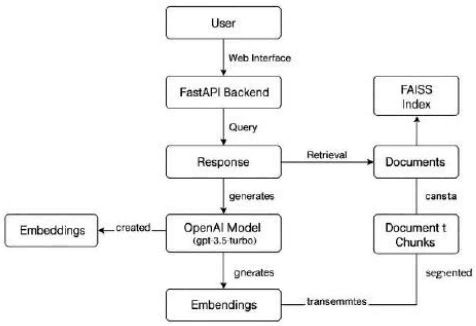

Fig 2. Basic Architecture

8. INTERFACE

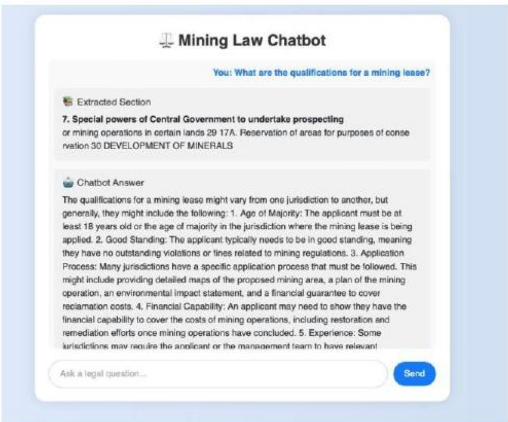

Fig 3. A Web Interface of the ChatBot

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

Plagiarism Report

Dupli Checker

Date: 15-05-2025

Plagiarism Scan Report

0% Plagiarism	0% Exact Match 0% Partial Match	100% Unique	

Words	644
Characters	4724
Sentences	66
Paragraphs	98
Read Time	4 minute(s)
Speak Time	5 minute(s)

Content Checked For Plagiarism

Abstract: This project based on the development of an AI powered chatbot is designed to help the mining industry of India.The system leverages Natural Language Processing NLP) and a vector similarity hunt medium to recoup applicable legal sections from authoritative documents similar as the Mines and Minerals(Development and Regulation) Act, 1957. Using OpenAI's GPT model as the language machine and FAISS for effective semantic reclamation, the chatbot enables druggies to interact through natural language queries and admit terse legal explanations predicated in sanctioned legal textbooks. This result democratizes access to nonsupervisory knowledge, minimizes legal nebulosity, and enhances decision-making across the mining assiduity. The system was tested using real nonsupervisory documents and demonstrates promising eventuality as a 24/7 accessible legal adjunct for mining drivers, 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 in natural language and receive context-aware responses grounded in official regulatory content. The

Sustainable Development Goals


16 PEACE, JUSTICE AND STRONG INSTITUTIONS
Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels

Pulication Certificates

CERTIFICATE
OF PUBLICATION

This is to certify that

ASHFAQ UR RAHMAN H N

has successfully published their research paper titled

AI-DRIVEN LEGAL CHATBOT FOR MINING REGULATIONS

in Prime Publication Hub

This certification recognizes the contribution of the author in advancing academic and scientific knowledge through their valuable research work. The paper has been peer-reviewed and meets the publication standards of Prime Publication Hub.

PPH

24120035	2412-0035
Author ID	**ISSN Number**

Issued By: Prime Publication Hub

Date of Issue: 16-05-2025

CERTIFICATE

OF PUBLICATION

This is to certify that

MOHAMMED AZEEM A

has successfully published their research paper titled

AI-DRIVEN LEGAL CHATBOT FOR MINING REGULATIONS

in Prime Publication Hub

This certification recognizes the contribution of the author in advancing academic and scientific knowledge through their valuable research work. The paper has been peer-reviewed and meets the publication standards of Prime Publication Hub.

PPH

24120035

Author ID

2412-0035

ISSN Number

Issued By: Prime Publication Hub

Date of Issue: 16-05-2025

CERTIFICATE

OF PUBLICATION

This is to certify that

SAGAR H N

has successfully published their research paper titled

AI-DRIVEN LEGAL CHATBOT FOR MINING REGULATIONS

in Prime Publication Hub

This certification recognizes the contribution of the author in advancing academic and scientific knowledge through their valuable research work. The paper has been peer-reviewed and meets the publication standards of Prime Publication Hub.

PPH

24120035	2412-0035
Author ID	**ISSN Number**

Issued By: Prime Publication Hub   **Date of Issue:** 16-05-2025