International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

MineMate – Your intelligent AI assistant for mining laws, regulations and industry insights

Sathvik N B Math
Department of Computer Science and Engineering
Presidency University
Bengaluru, India
sathviknbmath@gmail.com

Vijayeendra N
Department of Computer Science and Engineering
Presidency University
Bengaluru, India
swamyvijayeendra@gmail.com

Adithya H Hegde
Department of Computer Science and Engineering
Presidency University
Bengaluru, India
adithyahegde132@gmail.com

Pakruddin B
Department of Computer Science and Engineering
Presidency University
Bengaluru, India
fakrubasha@gmail.com

Abstract— **Simply, the digital governance accelerates citizen-government service interaction more with simple tools. In the paper, a chatbot will be developed using web scraping techniques dynamically collecting and updating needed information from government websites based on eGovernance. This will facilitate meaningful, intuitive conversations between users and the language model, Llama 3.2, making the access to public data and services even more interesting. All this will be collected in real time by the web scraper: knowledge base of the chatbot, including policies issued, deadlines covered by the government, citizen services offered, and many more. Inserting language comprehension and conversational capabilities of Llama 3.2 with gathering data in real time would be expected to create ease of interaction for the user, with minimum delay in information, and a realistic answer to citizens' inquiries. Discussions also include issues when web scraping on public websites is applied, legal perspectives, and efficiency in the integration of large language models into eGovernance systems. Results are readily available, responsive, and satisfactory concerning user satisfaction for AI-powered government service chatbots.**

Index Terms-Chatbot, eGovernance, Llama 3.2, Natural Language Processing

I. INTRODUCTION

Stakeholder engagement and information management have been transformed by artificial intelligence (AI) and natural language processing (NLP). Because legal papers are complicated and distributed, it can be difficult to ensure compliance with Acts, Rules, Regulations, and DGMS Circulars in highly regulated industries like mining [1]. The Coal Mines Act (1952), the Indian Explosives Act (1884), the Coal Mines Regulations (2017), and land acquisition laws such as CBA and LA [2] are among the legal frameworks that regulate the mining sector in India. Accessing and interpreting these regulations is still difficult, though, and frequently calls for manual searches, legal advice, or a thorough examination of the available documents [3]. The Indian mining industry is governed by multiple legal frameworks, including the Coal Mines Act (1952), Indian Explosives Act (1884), The Coal Mines Regulations (2017), and land acquisition laws like CBA and LA, making accessibility a persistent issue [2]. Even if these regulations are in place, it takes a lot of time and effort to find and comprehend pertinent legal information. The current approaches, which frequently result in decision-making delays and possible compliance issues, entail manual searches, legal consultations, or sorting through copious amounts of documentation [3]. To provide prompt and precise responses to industry stakeholders, an automated system that can interpret, retrieve, and display legal material in an understandable and structured manner is desperately needed.

This research proposes an AI-powered chatbot that utilizes NLP to provide real-time, context-aware regulatory information. Acting as a 24/7 digital compliance assistant, the chatbot will improve accessibility, enhance compliance efficiency, and minimize reliance on human legal experts. This study will explore the chatbot's design, implementation, and impact on legal accessibility, contributing to advancements in Regulatory Technology (RegTech) and AI-driven legal advisory systems [4].

II. LITERATURE REVIEW

Title - 01: Natural Language Processing in Chatbots:

Details: The fundamental methods of Natural Language Processing (NLP) that allow chatbots to comprehend and produce dialogues that resemble those of humans are examined in this study. Tokenization (dividing text into smaller pieces for analysis), named entity recognition (NER) (finding key terms like laws and regulations), and sentiment analysis (figuring out the purpose of queries) are some of the main techniques covered. The paper also emphasizes how machine learning models might help chatbots respond more intelligently over time. These methods enable chatbots to effectively automate responses across a range of domains, such as regulatory compliance and customer support. [5]

Drawbacks: The report highlights important issues such multilingual support issues, which make it difficult for chatbots to comprehend regional languages, and language ambiguity, which occurs when words with various meanings

confound the chatbot. Furthermore, chatbots managing sensitive user data raise data privacy issues, and integrating these systems into current platforms is challenging due to integration issues.

Title - 02: A Regulatory Chatbot for Mining Industries

Details: This study introduces a chatbot created especially to help mining professionals by offering prompt answers to legal and regulatory questions. It retrieves pertinent mining regulations and compliance requirements using a rule-based retrieval mechanism and sophisticated natural language processing techniques. The chatbot improves operational efficiency and guarantees regulatory compliance in the mining industry by automating the handling of legal queries. The ability of the chatbot to adjust to frequent legislative changes through ongoing updates is also covered in the study. [6]

Drawbacks: Complex legal terminology is one of the main obstacles, making it hard for chatbots to understand and react appropriately. In order to ensure accuracy, regulatory modifications also necessitate regular data refreshes. Another major obstacle to widespread adoption of big legal databases is the high memory and computing expenses involved.

Title - 03: Chatbot to Respond to Text Queries Pertaining to Various Acts, Rules

Details: The goal of this project is to create a chatbot that combines cutting-edge AI technologies to effectively extract and process legal requirements. The following are included in the methodology:

- FAISS and LangChain: to improve the speed of document retrieval and search.
- Hugging Face AI models: to increase answer accuracy and query comprehension.
- Llama 2 7B Chat is a potent generative AI model that provides intuitive answers to legal questions.

By combining these technologies, the chatbot makes mining laws more accessible and enables professionals to get immediate legal advice. [7]

Drawbacks: Complex legal terminology is one of the main obstacles, making it hard for chatbots to understand and react appropriately. In order to ensure accuracy, regulatory modifications also necessitate regular data refreshes. Another major obstacle to widespread adoption of big legal databases is the high memory and computing expenses involved.

Title - 04: MineBot: Chatbot to Respond to Text Queries Pertaining to Mining Regulations

Details: MineBot uses external data sources to improve response accuracy in its role as a legal assistant for mining professionals. In contrast to conventional rule-based frameworks, MineBot incorporates:

1. Wikipedia & Google API – for fetching real-time legal references.

2. Custom legal databases – for industry-specific mining laws.

Because of this combination, MineBot can offer thorough and current legal responses, which makes it an invaluable tool for compliance teams. [8]

Drawbacks: Due to its reliance on external APIs, the chatbot may experience latency problems and data accuracy issues if its sources are out-of-date. Furthermore, legal databases need to be updated frequently in order to maintain their reliability because they are extensive and dynamic.

Title - 05: Chatbot for Coal Mining Industry

Details: This is an AI-powered chatbot created especially to help coal mining professionals by giving them immediate access to environmental and regulatory requirements is presented in this study. By providing regulatory assistance and processing legal queries, the chatbot simplifies compliance operations. It increases productivity in the coal mining sector by cutting down on the amount of time spent on legal research. [9]

Drawbacks: The complexity of legal language, which makes NLP-based interpretation challenging, the high development costs resulting from the requirement for large-scale AI models, and ongoing maintenance to guarantee legislation stay current are the primary obstacles.

Title - 06: MineShark AI: Respond to Text Queries and Regulations Applicable to Mining Industries

Details: MineShark AI brings Generative AI (Gen AI) to the legal chatbot space, enabling the creation of dynamic answers to regulatory questions pertaining to mining. Since MineShark AI produces answers in real time, unlike rule-based bots, it is extremely flexible when it comes to answering all kinds of legal questions. This method expedites the verification of compliance and enhances accessibility. [10]

Drawbacks: Ensuring the accuracy of AI-generated solutions, resolving ethical issues with AI in legal decision-making, and keeping up with regular revisions to reflect new legislation are the main problems. Furthermore, there are worries about legal terms being misunderstood when AI-generated content is relied upon.

III. METHODOLOGY

Large Language Models represent complex artificial intelligence capable of understanding, generating, and manipulating human languages. They are based on deep learning architectures that generally involve or are built around the Transformer model, which allows them to process substantial amounts of text in learning patterns, context, and even linguistic structures. LLMs go one step further in learning the meanings of individual words, and also relationships among words, sentences, and broader textual elements that will permit them to construct coherent, contextually relevant responses.

They work by embedding layers that break down text into smaller units-known as tokens-and turn them into vectors. These vectors feed into multiple layers within the model, which include self-attention mechanisms for weighting the importance of a token in relation to others for contextual captures. Fully capturing the nuanced relationships found within these layers enables refinements to be made by the deeper layers of the model.

They get pre-trained on general text data and then fine-tuned to do translation tasks, summarization, or conversation. Along the way, they learn grammar, facts, reasoning, and even subtleties like tone and intent. Large language models like Llama 3.2 8B have great understandability and the capability to generate human-like text, making them very effective for chatbots, content creation, and automatic question answering applications [3]. They could handle complex representational language tasks, including multi-lingual and domain-specific interactions, which are very suitable for use cases in eGovernance, customer services, and education.

A. Llama 3.2 8B

The newest, large language model released by the company is from Meta, formerly Facebook. Llama 3.2 8B broadens on the already published versions of Llama because it extends capabilities to better understand and generate text that closely resembles the form and structure of human language across a broad range of subjects. Like other LLMs, it is Transformer architecture-based, but Llama 3.2 8B is optimized for efficiency, so it can perform quite perfectly even on fewer computational resources than some of the large models similar to GPT-4. Llama 3.2 8B was subjected to the strongest training with a large, multilingual and diversified source dataset [3]. Therefore, it exhibits varying skills of performance in respect to different languages and domains of specialty. It enables flawless work in complicated situations-for example, legal document processing, scientific research, etc. This also comes along with fine-tuning enhancements which have made it easier to adapt with minimal data for special tasks. Llama 3.2 8B has adapted features on reasoning, summarization, and even sentiment analysis, making it capable of doing many more things than other massive language models.

B. Llama 3.2 8B vs LLMs

B.1. Llama 3.2 8B vs GPT-4

Some of the most advanced LLMs developed in this AI landscape so far include Llama 3.2 8B and GPT-4, designed with highly sophisticated deep penetration natural language processing tasks at their core. Its focus on optimal utilization of resources also allows it to do so much better than stronger demands made by computations and is also extremely versatile for deployment at a large scale, including chatbots and real-time support in content creation provided by eGovernance services.

Comparatively, GPT-4, has been reported to possess wide scalability and generalization capabilities. Having been trained on a much bigger corpus than the previous models, GPT-4 is significantly more versatile in an enormously wider spectrum of tasks ranging from creative writing, reasoning, and coding. That makes the strength of GPT-4 on its greater capacity for nuances in the responses and the depth it could exhibit while holding context in multi-turn conversations.[4]

Llama 3 405B performs approximately on par with GPT-4 (0125 API version). It outperforms GPT-4 in multiturn reasoning and coding tasks. However, it underperforms compared to GPT-4 in multilingual prompts (Hindi, Spanish, and Portuguese). [3]

B.2. Llama 3.2 8B vs Gemini 1.5

Llama 3.2 8B and Gemini AI are two releases of very high-functioning large language models with different explicit design advantages for some specific uses in AI. In contrast, Gemini AI by DeepMind is a multi-modal AI model that brings language processing together with a huge number of kinds of data - images, videos, and so on. So, Gemini AI may be used for solving impressively diverse tasks, moving far beyond the strict text-based interaction in order to solve challenging problems in such spheres as healthcare and autonomous systems to new types of content. Its multimodality makes it rather distinct from Llama 3.2 8B, which is still much more text- and conversation-oriented. Of course, Llama 3.2 8B may perform exceptionally well with tasks of real-time natural language understanding and conversational dynamics, but one of the strengths of the Gemini AI technology is its ability to fuse different kinds of data toward ever more holistic and contextual problem solving.

While, Gemini 1.5 Pro achieves 100% recall up to 530k tokens and >99.7% recall up to 1M tokens [5]. Llama 3.2 8B models demonstrate perfect needle retrieval performance, successfully retrieving 100% of needles at all document depths and context lengths. We also measure performance on Multi-needle (Table 21), a variation of Needle-in-a-Haystack, where we insert four needles in the context and test if a model can retrieve two of them. Our Llama 3 models achieve near perfect retrieval results. [3]

B.1. Llama 3.2 8B vs Gemma 2

Among the newer state-of-the-art language models are Llama 3.2 8B and Gemma 2, built for various features of different use cases in the artificial intelligence landscape. The Llama 3.2 8B architectural design intends to produce context-aware responses in a wide scope of languages while carrying flexibility with fine-tuning for specific domains, making it well suited for very specialized tasks. While much larger in scope, Gemma AI will have much more complex data processing applications aside from its capabilities in natural language [6]. Llama 3.2 8B, is basically an exercise in conversational ability, but Gemma AI would have far more robust capabilities with multitasking applications, handling complex analytics, and making it even stronger in applications such as enterprise business intelligence, predictive modelling, and many other things. While Llama 3.2 8B is fitted for effective and smart, intelligent natural conversations, the entire solution from Gemma AI integrates natural language processing with superior data interpretation to create an expertly crafted product-ideal for organizations looking for deep AI solutions.[6]

Fig. 1. Human evaluation of Llama 3.2 8B

Model	MGSM	Multilingual MMLU
Llama 3 8B	**68.9**	**58.6**
Mistral 7B	29.9	46.8
Gemma 2 9B	53.2	–
Llama 3 70B	**86.9**	**78.2**
GPT-3.5 Turbo	51.4	58.8
Mixtral 8×22B	71.1	64.3
Llama 3 405B	**91.6**	83.2
GPT-4	85.9	80.2
GPT-4o	90.5	**85.5**
Claude 3.5 Sonnet	**91.6**	–

Fig. 2. Llama Multi Lingual test scores

Model	MGSM	Multilingual MMLU
Llama 3 8B	**68.9**	**58.6**
Mistral 7B	29.9	46.8
Gemma 2 9B	53.2	–
Llama 3 70B	**86.9**	**78.2**
GPT-3.5 Turbo	51.4	58.8
Mixtral 8×22B	71.1	64.3
Llama 3 405B	**91.6**	83.2
GPT-4	85.9	80.2
GPT-4o	90.5	**85.5**
Claude 3.5 Sonnet	**91.6**	–

Fig. 3. Proeficieny test scores of Llama 3.2 8B

C. Working

The web-based chatbot combines the Llama 3.2 8B language model with a React.js frontend, a Flask backend, and MySQL for database management. The application's main goal is to provide interactive secure conversational services. Through the online interface, users can interact and submit and process inquiries. User information, including name, email, and hashed password, is safely saved upon registration. SMTP protocols are used to send confirmation emails, which strengthen user verification and authentication [12].

User questions are sent to the Llama 3.2 model via the backend, which interprets natural language input and provides precise, context-aware answers. Although its primary optimization is for English, the model accommodates multilingual input and has efficient inference capabilities thanks to improvements for 8B parameters [11].

The Llama 3.2 8B model has performed exceptionally well on common benchmarks, including:

- **MMLU (Massive Multitask Language Understanding)** for general reasoning and knowledge tasks.
- **HumanEval** for code generation and logical problem-solving.
- **ARC and TruthfulQA** for evaluating common-sense reasoning and factual accuracy [13].

According to research, Llama 3 performs better than previous iterations in activities including following instructions and in sophisticated applications like chatbots and code creation aids. [11] [14].

With user-level authentication, password hashing, and encrypted data, this design offers strong security in addition to functionality. These architectural techniques are in line with current guidelines for user data protection and web security [12].

For AI-powered web applications, the combination of frontends like React with LLMs like Llama provides quick deployment, adaptability, and maintainability. The model can be executed locally (with Massive Multitask Language Understanding, or MMLU, is used for knowledge and general reasoning problems.

HumanEval for logical problem-solving and code generation.

ARC and TruthfulQA for assessing factual accuracy and commonsense reasoning [13].

According to research, Llama 3 performs better than previous iterations in activities including following instructions and in sophisticated applications like chatbots and code creation aids.

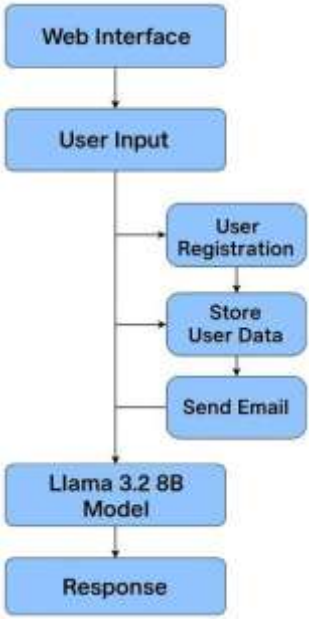

Fig. 4. Working of the chatbot

1. **Web Interface (React.js)** – Users interact with the chatbot via the frontend.
2. **User Registration/Login** – User data is securely registered, with:
   - **Password hashing** (bcrypt or similar)
   - **Storage in SQL database**
   - **Email confirmation** using SMTP
3. **User Input** – Message from user sent via API.
4. **Backend Server (Python/Flask)** – Receives and processes input.
5. **LLaMA 3.2 8B** – The model is queried for relevant responses.
6. **Response Returned** – LLM-generated answer is sent back to the user.

IV. Experimental Results

The results from our chatbot experiments show that it effectively answers user questions by pinpointing and retrieving relevant loans and schemes tailored to different sectors. When a query is received, the chatbot actively searches through a wide range of resources, including PDF documents, to pull out specific information related to the sector in question—be it agriculture, education, small businesses, or other focused areas. The responses from the chatbot are customized to match user intent, offering clear yet thorough details about available government schemes or loan options, eligibility requirements, and potential benefits in each sector. This ability to provide sector-specific responses emphasizes the chatbot's role as a valuable eGovernance tool, allowing users to quickly find relevant information without the need to navigate through complicated documentation. The findings highlight the system's accuracy, responsiveness, and effectiveness in improving user interaction and satisfaction.

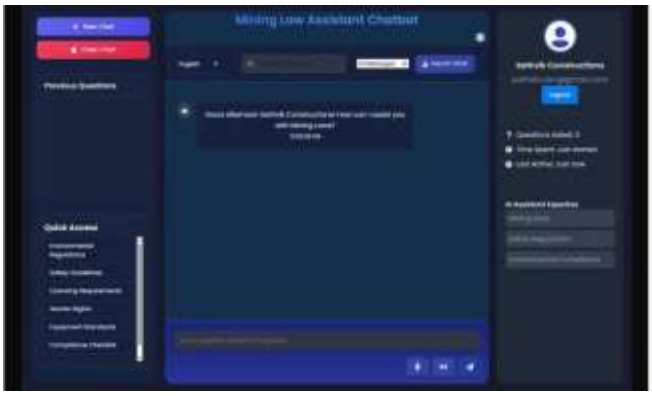

Fig. 5. UI/UX of the chatbot

V. Conclusion and Future Work

Chatbot technology, which combines Artificial Intelligence (AI) and Natural Language Processing (NLP), provides a revolutionary way to solve the persistent problems with obtaining and understanding mining industry laws. The creation of a chatbot that can comprehend user enquiries and extract pertinent data from intricate legal documents, including the Indian Explosives Act (1884), the Coal Mines Act (1952), and numerous DGMS circulars, can greatly expedite operational decision-making and regulatory compliance.

By acting as a centralised, round-the-clock digital assistant, this chatbot lessens the need for legal professionals to handle routine questions. It makes complex legal jargon easier to understand for stakeholders and field workers, and it encourages prompt adherence to compliance standards. Its applicability to practical problems that go beyond technical mining operations is further increased by include acts pertaining to land acquisition and rehabilitation.

The suggested system establishes the foundation for an intelligent, user-centric Regulatory Tech (RegTech) tool designed for the mining industry, despite ongoing issues with data accuracy, model maintenance, and legal validation. Such AI-driven technologies are essential for developing more intelligent, secure, and transparent regulatory environments in mining and other highly regulated industries as digital revolution continues to change industrial processes.

REFERENCES

[1] Jurafsky, D., & Martin, J. H. (2021). Speech and Language Processing. Pearson.

[2] Chui, M., Manyika, J., & Miremadi, M. (2018). AI adoption advances in the legal sector. McKinsey Global Institute..

[3] Directorate General of Mines Safety (DGMS). (2017). The Coal Mines Regulations, 2017. Government of India.

[4] Surden, H. (2020). "Artificial Intelligence and Law: An Overview. Harvard Journal of Law & Technology, 33(1), 1-30.

[5] Natural Language Processing in Chatbots: A Review.ResearchGate, 2023. ResearchGate

[6] A Regulatory Chatbot for Mining Industries. MDPI, JETIR, IRJMETS, 2023.

[7] C Chatbot to Respond to Text Queries Pertaining to Various Acts, Rules. JETIR, 2023.

[8] MineBot: Chatbot to Respond to Text Queries Pertaining to Mining Regulations. IJRPR, 2024.

[9] Chatbot: For Coal Mining Industry. PhilArchive, IRJMETS, IJRPR, 2023.

[10] MineShark AI: Respond to Text Queries and Regulations Applicable to Mining Industries. JETIR, PhilArchive, 2025.

[11] Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T. & Jegou, H. (2024). Llama 3: Open Foundation and Instruction-Tuned Models. arXiv:2407.21783 Gwennap, L., 2020. Groq rocks neural networks. Microprocessor Report, Tech. Rep., jan.

[12] Mohan, R., & Ragothaman, A. (2023). A Framework for Secure User Authentication in AI Applications. International Journal of Computer Applications, 186(25), 1-7.

[13] Meta AI Research (2024). The Llama 3 Herd of Models: Capabilities, Benchmarks, and Optimizations.

[14] Yang, Y., Gao, L., & Lin, Z. (2024). Context Expansion in Llama 3 for Long Document Understanding. arXiv:2404.19553