Chatbot to respond to text queries pertaining to various Acts, Rules, and Regulations applicable to Mining Industries.

Nithya B A
Assistant Professor,
nithyaba3@gmail.com

GUBA KUSHAL NAIDU
UG Student,
kushalnaidug26@gmail.com

S. SANTHOSH
UG Student,
Santhoshseelam@gmail.com

K. Mohan Reddy
UG Student,
Nani18112003@gmail.com

G. Rithvik
UG Student,
rithvikgangari@gmail.com

Dept of CSE, Presidency University, Bengaluru, India

ABSTRACT:

Mining rules in India are composed and are spread in many actions, rules and modifications, making it difficult for stakeholders - especially for workers and junior officers in the region - to reach accurate information. This research introduces an online, multilingual chatbot that can answer regulatory questions in English, Hindi, Telugu and Kannada. By using standard web technologies and integrating a mild backend NLP engine, Chatbot provides the reference UPs that are legal reactions and improves access to legal knowledge related to mining. The system supports multilingual input and adjusts dynamic reactions to the user interface and user preferences. By automating query management, it reduces the dependence on human legal advisors and increases properly conformity efficiency.

1. INTRODUCTION

In Mining operations should follow a wide range of legal frameworks, including minerals and minerals (development and regulation), mineral license rules and various state -specific changes. Understanding and implementing these rules requires sufficient legal literacy – often a barrier to field engineers, supervisors and operators in rural or multilingual areas. Traditional solutions such as printed manuals, PDF or legal consultations in tradition are either inaccessible or disabled in real-time fields.

In response, the project proposes a domain-specific Chatbot that can interpret user issues on mining rules and provide accurate response to many languages. The purpose of the system is to be simple, sharp and widely available. Multilingualism is a main function, which reflects the linguistic diversity of mining areas in India. By streamlining access to legal information, Chatbot helps users make informed decisions, reduce legal risk and promote safe, more obedient mining tasks.

2. LITERATURE REVIEW:

The integration of AI into the legal domain is well documented, where early systems such as Micin (for medical region) pave the way for specialist systems in the law. Later, knowledge -based systems such as legal specialist systems (read) were used to coded logical rules from case law. However, these systems are struggling with real world language variability and essential intensive domain modeling.

The recent NLP progress, especially the transformer- based models such as Burt, GPT and their multilingual variants (e.g., mBERT, XLM-R), is to understand the natural language dramatic machine. Chatbots like LawDroid and DoNotPay explain how AI can be used on legal services. However, most of the implementation focuses on cases of western legal systems and the use of English -language.

In the Indian context, research is still emerging. There are some attempts at Indian language chatbot, often in bank or customer service. Legal language translation is a challenge due to the vocabulary's ambiguity and lack of structured data sets. Our project contributes with a new approach to combining rule-based recovery with multilingual support to fit the Indian mining regulator environment.

3. METHODOLOGIES:

The system follows a modular architecture composed of five layers: user interface, session administration, query preparedness, multilingual changes and response generations. Here's a breakdown:

3.1 Front Design

Frontend was developed using HTML, CSS and Jinja template. Five main pages used:

Login.html: Certification users and shows notice of unsuccessful efforts.

Register.html: JavaScript handles new user registration with information.

Select_language.html: Offers a drop -down menu to select one of the four languages.

Chatbot.html: The most important chat interface that adapts entrance holders and button text based on the selected language.

Base.html: A shared layout containing styles, header and script references for all sides.

3.2 Backend Flow

Flask handles routing and user session administration. Post requests are used to process the form and handle the chatbot query. The selected language is stored in the session and is used to customize both frontal reproduction and backend response formatting.

3.3 Language Personalization

Each user request is marked with a language prefix. This leads the Backend NLP engine to use appropriate translation rules and select from the language-specific QA database. Language-specific sentences, formal structures and vocabulary are preserved using customized translation words.

4. DATACOLLECTION & PREPROCESSING:

Acquiring and preparing data underpinning our chatbot's functionality is a crucial stage of our research methodology. The mining industry, by nature, generates a wealth of documentation encapsulating the rules, regulations, and guidelines that govern its operations. These documents, often voluminous and diverse, are predominantly distributed in the ubiquitous Portable Document Format (PDF). To render them usable by our chatbot, we embark on an intricate data collection and preprocessing journey.

4.1 Legal Data Sources

Data was sourced from:

- Indian Bureau of Mines documentation
- Ministry of Mines website
- State mining department circulars
- Open-access mining legal guides
- Both Acts (e.g., MMDR Act, 1957) and Rules (e.g., MCR 2016) were collected in machine-readable formats (PDF, HTML, DOCX).

4.2. Preprocessing Pipeline

Text Extraction: OCR was implemented where documents were scanned. Tools such as Tesseract and PDF Miner were used.

Translation: Microsoft translator and Google translation were originally used, followed by manual improvements to preserve legal accuracy.

Tokenization: The sentences were divided and marked as questions or reactions.

Embedding: Sentence transformer (for example, all- MiniLM-L6-v2) was used to convert the text of the vector that involves semantic comparison.

Storage: QA pairs were stored in the SQLITE database indexed by the language.

This pipeline secured the same structure of language and adapted to the recovery accuracy.

5. CHATBOT DEVELOPMENTS:

The development of our chatbot for mining industry rules and regulations is a meticulously planned and executed process, bringing together a constellation of technologies and components that work in harmony to offer a seamless and user-friendly experience. The chatbot development not only encompasses the technical aspects of programming and natural language processing but also prioritizes the user's interaction with mining regulations, enhancing accessibility and understanding.

The chatbot was built using a hybrid approach:

5.1 Ruler -based arguments

Newly ordered keyword triggers were mapped for reactions, especially for frequent questions such as "Royalty Payment", "Lease Renewal" and "Illegal mining”.

5.2 Semantic Matching

For the query for free shape, built-in was used to match the entrance with the existing QA pairs. With a threshold point of 0.70 to filter the results of low confidence, Cosine Like position was used to rank matches.

5.3 Language customization

Based on selected languages:

- The entrance issues were interpreted through language -specific tools.
- The answer was restored from a language-specific QA set.
- Dynamically adjusted for translation provides templates to the marked and sections.

5.4 Fallback Mechanism

For unknown questions, the system suggested predetermined subjects (FAQ) or recommended contacting the mining department. This ensured user storage without providing misleading answers.

6. RESULTS:

The evaluation of our chatbot's performance in assisting users with mining rules and regulations reveals a series of noteworthy findings, shedding light on its effectiveness in addressing the challenges associated with navigating this complex landscape. Here, we present a detailed breakdown of the results, including key metrics and user feedback.

The system was evaluated across three dimensions:

6.1 Functional Testing

User authentication: Login/Registration for 100+ test accounts.

Language change: Seamless interface adaptation in four languages.

Chat flow: Treated successfully 160+ legal questions over 2's reaction time.

6.2 Accuracy Metrics

Metric | Price |

Query Recognition Accuracy | 85.2% |

Semantic Match Accuracy | 88.6% |

Language translation accuracy | 91.3% (manually confirmed) |

6.3 Feedback from the user

Pilot samples revealed by 30 users (the author's trainee, mining engineers):

- 92% found Chatbot useful for answering basic legal questions.
- 87% considered as a language experience as "very good”.
- 74% preferred it on PDF manual or internal FAQ.

7. DISCUSSIONS:

The study reflects the viability of the creation of a domain -specific, multilingual legal chatbot using light technologies and open-source tools. The most important achievement is its access - users can achieve regulatory clarity in seconds without legal expertise or English skills.

However, challenges remain:

7.1 Legal language translation

Legal expressions are not always clean. There is no direct equivalent to certain conditions (e.g. "concession," recycling "). Manual review is crucial to ensuring legal loyalty, especially in official or used contexts.

7.2 Relevant ambiguity

Users can ask unclear questions (e.g. "what is the rule for mining?"). Without deep semantic analysis depends on the chatbot key word of proximity and can give a normal answer.

7.3 Scalability

Adding new actions or supporting multiple languages (e.g. Marathi, Tamil) needs to expand datasets and translations, which is resource intensive. Integrating the AI model as GPT-4 can increase the short depth but introduces problems with cost and regulation.

Despite these boundaries, the system has shown capacity for institutional deployment as a decision-supporting equipment in field operations, mining inspection and training programs for institutional distribution.

8. REFRENCES:

1. Ministry of Mines, Government of India. Mines and Minerals (Development and Regulation) Act, 1957. https://mines.gov.in

2. Indian Bureau of Mines. Mineral Concession Rules, 1960 & Mineral Conservation and Development Rules, 2017. https://ibm.gov.in

3. Patel, N., & Shah, A. (2020). "AI-Based Legal Chatbot for Indian Judiciary System." International Journal of Innovative Technology and Exploring Engineering, 9(4), 1402–1406.

4. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." arXiv preprint arXiv:1810.04805.

5. Conneau, A., et al. (2020). "Unsupervised Cross- lingual Representation Learning at Scale." Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL).

6. Chavan, V., & Mane, S. (2022). "Survey on Multilingual Chatbots: Techniques and Challenges." International Research Journal of Engineering and Technology (IRJET), 9(1), 125–130.

7. DoNotPay. The World's First Robot Lawyer. https://donotpay.com

8. LawDroid. AI Legal Assistant for Lawyers and Clients. https://www.lawdroid.com

9. Microsoft Translator. AI-Powered Translation Services. https://www.microsoft.com/translator

10. Google Cloud Translation API. https://cloud.google.com/translate

11. Tesseract OCR Engine. https://github.com/tesseract-ocr

12. Reimers, N., & Gurevych, I. (2019). "Sentence- BERT: Sentence Embeddings using Siamese BERT- Networks." Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP).

13. Akash Takyar. HOW TO BUILD AN AI-POWERED CHATBOT? https://www.leewayhertz.com/ai-chatbots/

14. Khanna, A., Pandey, B., Vashishta, K., Kalia, K., Bhale, P., Das, T.: A study of today's A.I. through chatbots and rediscovery of machine intelligence. Int. J. u-e-Serv. Sci. Technol. 8, 277–284 (2015). https://doi.org/10.14257/ijunesst.2015.8.7.28 .

15. Klopfenstein, L., Delpriori, S., Malatini, S., Bogliolo, A. *The rise of bots: a survey of conversational interfaces, patterns, and paradigms.* In: Proceedings of the 2017 Conference on Designing Interactive Systems, pp. 555–565. Association for Computing Machinery (2017).

16. Suarez-Tangil, G., Tapiador, J.E., Peris-Lopez, P., Ribagorda, A. *Evolution, Detection and Analysis of Malware for Smart Devices.* IEEE Commun. Surv. Tutor. 2014, 16, 961–987. [CrossRef]

17. Jurafsky, D., Martin, J.H. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition,* 2nd ed.; Prentice Hall PTR: Upper Saddle River, NJ, USA, 2007.

18. Bavaresco, R., Silveira, D., Reis, E., Barbosa, J., Righi, R., Costa, C., Antunes, R., Gomes, M., Gatti, C., Vanzin, M., et al. *Conversational agents in business: A systematic literature review and future research directions.* Computer. Sci. Rev. 2020, 36, 100239. [CrossRef]

19. Sun, Y., Zhang, Y., Chen, Y., Jin, R. *Conversational Recommendation System with Unsupervised Learning.* In Proceedings of the 10th ACM Conference on Recommender Systems, Boston, MA, USA, 15–19 September 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 397–398. [CrossRef]

20. Jusoh, S. *Intelligent Conversational Agent for Online Sales.* In Proceedings of the 2018 10th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Iasi,

Romania, 28–30 June 2018; pp. 1–4. [CrossRef]

21. Majumder, A., Pande, A., Vonteru, K., Gangwar, A., Maji, S., Bhatia, P., Goyal, P. *Automated Assistance in E-commerce: An Approach Based on Category-Sensitive Retrieval.* In Advances in Information Retrieval; Springer: Cham, Switzerland, 2018; pp. 604–610. [CrossRef]

22. Koetter, F., Blohm, M., Kochanowski, M., Goetzer, J., Graziotin, D., Wagner, S. *Motivations, Classification and Model Trial of Conversational Agents for Insurance Companies.* In Proceedings of the 11th International Conference on Agents and Artificial Intelligence—Volume 1: ICAART, INSTICC, Prague, Czech Republic, 19–21 February 2019; SciTePress: Prague, Czech Republic, 2019; pp. 19–30. [CrossRef]

23. Arjun Sha (July 29, 2023). *How to Train an AI Chatbot with Custom Knowledge Base Using ChatGPT API.* https://beebom.com/how-train-ai-chatbot-custom-knowledge-base-chatgpt-api/

24. Sweety Sahani & Sushmitha Mary (May 21, 2022). *Chatbot Using Python.* https://www.ijraset.com/research-paper/chatbot-using-python

25. Langchain (18 February, 2024). *Langchain: PDF Chat App (GUI) | ChatGPT for Your PDF FILES | Step-by-Step Tutorial.* https://www.youtube.com/watch?v=RIWbalZ7sTo