



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

Vinamra Khandelwal , Param Dhingana, Shruti Lalwani , Shivesh Mishra , Sudeep Ganguly , Prof Dipti Chauhan

Member , Member , Member , Member , Member , Guide

Department of Artificial Intelligence and Data Science

Prestige Institute of Engineering Management and Research(PIEMR), Indore, India

Abstract : The mining industry is characterized by an intricate web of rules and regulation, essential for ensuring safety environmental compliance, and operational efficiency. In this research, we address the challenge of facilitating access to, and understanding of mining rules and regulations through the development of a chatbot. We employ state-of-the-art- technologies, including Langchain, FAISS(CPU), Hugging Face, Python, and Llama 2 7B Chat, to create a comprehensive solution. Our approach involves a multi-step procedure, encompassing chunks, adding these chunks to a vector index, and efficiently retrieving them when users pose a question. The chatbot not only streamlines the retrieval of relevant information but also generates hypothetical answers to user queries. Key findings of our research demonstrate that chatbot significantly enhances accessibility with this vital information. The significance of this research lies in its potential to improve compliance and safety in the mining industry, ultimately contributing to more responsible and sustainable mining practice.

Keywords: Mining Regulations, Chatbot, Langchain, FAISS(CPU), Hugging Face, Python, Llama 2 7B Chat, Management Information System, Information Retrieval

I. INTRODUCTION

In today's fast-paced world, where time is a precious commodity, texting has emerged as one of the most common forms of communication. Hence, chatbots are becoming a crucial part of businesses' operations, regardless of their size or domain. The concept of chatbots can be traced back to the idea of intelligent robots introduced by Alan Turing in the 1950s. ELIZA was the first chatbot developed by MIT professor Joseph Weizenbaum in the 1960s. Since then, AI-based chatbots have been a major talking point and a valuable tool for businesses to ensure effective customer interactions.[1]

The mining industry, a critical pillar of global resource extraction, is underpinned by an intricate web of rules and regulations aimed at ensuring safety, environmental sustainability, and operational efficiency. Compliance with these regulations is paramount, yet the complexity and sheer volume of mining rules and guidelines often pose a significant challenge for industry professionals and stakeholders. In response to this challenge, we have embarked on a research endeavor that seeks to revolutionize how individuals interact with and comprehend mining regulations [12].

Our approach involves the creation of a cutting-edge chatbot harnessed with the power of advanced technologies such as Langchain, FAISS (CPU), Hugging Face, and Llama 2 7B Chat, all thoughtfully integrated into a cohesive system. This research aims to provide a holistic solution to the multifaceted problems associated with mining regulations by facilitating easier access, comprehension, and interaction with these essential guidelines. Our contribution is underpinned by a meticulously structured workflow involving the loading of PDF documents, the segmentation of content into manageable portions, the establishment of a vector index for efficient data retrieval, and the innovative generation of hypothetical answers to user queries [2].

In doing so, our chatbot not only simplifies the retrieval of relevant information but also fosters a user-friendly environment for understanding and navigating the labyrinth of mining rules and regulations.

The significance of this research transcends the immediate domain of technology; it holds the potential to enhance compliance, safety, and operational excellence within the mining sector, thereby contributing to more responsible and sustainable mining practices, a goal that resonates deeply with the evolving landscape of environmental stewardship and resource management in the 21st century. Chatbots at an institutional level will act as counselors and advisors for hundreds of students simultaneously,

employing the same effective approach and regularly updating various guidelines and crucial data. The world's leading organizations in digital realms have launched their own chatbots at various intervals.

Year	Timeline of Chatbots
1950	Chatbots Revolution Concept of truly intelligent Machine
1966	Eliza – MIT – Simulate Human Conversation
1972	Parry- Added Conversational Strategy.
1988	Parry- Added Conversational Strategy.
1992	Dr. SBAITSO- Speech Synthesis Program
1995	Alice- Artificial Linguistic Internet Computer Entity – Heuristic Patten
2001	SMARTERCHILD – Fun Personalized network ; Precursor to Apple's SIRI
2006	IBM's WATSON – Natural Language Processing; Machine Language.
2010	SIRI – Apple's IOS, Natural Language UI
2012	Google Now –uses natural language for google search on mobile
2015	Alexa – Amazon Echo Device; using language processing Algorithms
2015	CORTANA- Bing Search; Natural Voice; Different Language
2016	Facebook user bots
2016	TAY – Microsoft to mimic the speech and habit of teenage girl
2017	Google Assistant – Virtual Assistant
2018	LaMDA – Large Language Model from Google AI
2022	ChatGPT – Large Language Model by Open AI
2023	Bard – Large Language Model from Google AI

2. LITERATURE REVIEW:

The first known chatbot was Eliza, developed in 1966, whose purpose was to act as a psychotherapist returning the user utterances in a question form [4]. Alan Turing in 1950 proposed the Turing Test (“Can machines think?”), and it was at that time that the idea of a chatbot was popularized [5]. It used simple pattern matching [6] and a template-based response mechanism. Its conversational ability was not good, but it was enough to confuse people at a time when they were not used to interacting with computers and give them the impetus to start developing other chatbots [7].

Along with the improvements made, barriers to information access have declined. For Information and Communication Technology (ICT). Electronic devices continue to be more powerful [8]. Meanwhile, due to the development of wireless protocols and technologies, Connections to the Internet are no longer tied to time and space. The Fusion of Faster network, efficient applications, better operating system, better user interface, Enhanced performance technology, and expanded ecosystem of application markets made modern smart phones successful [9]. Effortless access, process, and production of information are now possible using smart phones. After the widespread use of smart phones, the development of messaging applications. This has resulted in conversational agents, also called chatbots in the industry, becoming one Trending solutions for trivial problems. Conversational agents are software applications that Interact with users for different purposes using conversation [10]. Chatbots are widely Used in various fields such as business, medical and health care fields and disaster management Bavaresco et al. [11] Reviewed recent studies and found that chatbots are the most common.

Explored in commerce domain covering sub-sectors such as e-commerce, sales, shopping, and flight bookings. Sun et al. [12] Adopted chatbots for personalized recommendation systems to build a virtual sales agent using deep learning technologies. Jusoh [13].

Proposed a method to persuade e-customers to buy products through recommendation and Conversation. [14] Created a chatbot to help category-sensitive customers. Retrieved technology and quality improvement. koeter et al. [15]

3. METHODOLOGIES:

The development of our chatbot for mining industry rules and regulations is guided by a systematic approach. We combine the capabilities of Langchain, FAISS (CPU), Hugging Face, and Llama 2 7B Chat to create an integrated system [25]. The primary workflow involves loading PDF files containing mining regulations [26], splitting the text into manageable chunks, adding these

chunks to a vector index, and establishing a mechanism for efficient retrieval. Additionally, the chatbot is designed to generate hypothetical answers based on the retrieved chunks, thus enhancing user interactions with the regulations.

An AI-based chatbot is more complex as they have continuous learning capability, they can communicate like a normal person, it knows how to categorize information, how to store information, and along with this based on the previous history how to help in a more proactive way to the user.

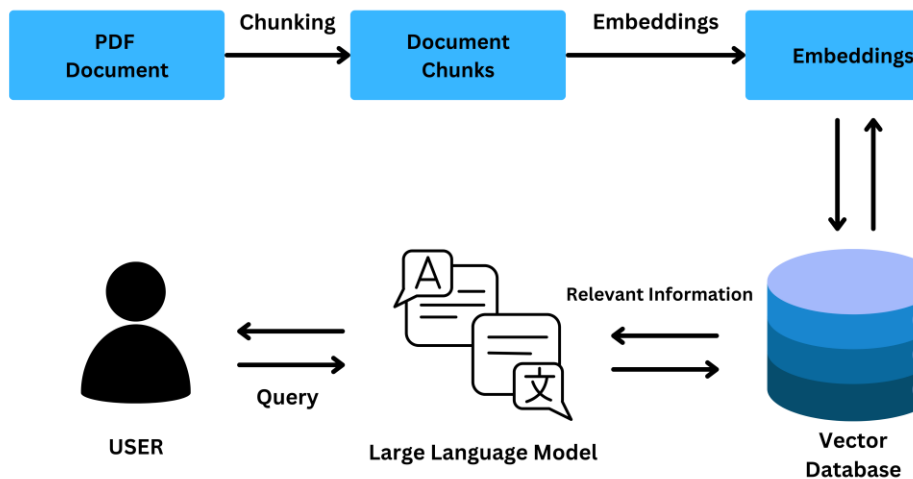


Fig.1.Chatbot Conversation Framework

The communication is completed with the responses from the conversation initiated on either retrieval or generative basis resulting in the closed or open domain. The Fig.1 explains the possible solution and complexities on both generative and retrieval based communication (conversation and response) in the closed and open domain.

4. DATA COLLECTION AND PREPROCESSING:

Acquiring and preparing the data underpinning our chatbot's functionality is a crucial stage of our research methodology [25]. 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 Data Collection: Our data collection process involves the systematic acquisition of PDF documents containing mining rules and regulations. These documents are sourced from a multitude of reputable industry and governmental sources [23]. The diversity and origin of these documents ensure a comprehensive coverage of the mining sector's regulatory landscape, accounting for both regional and industry-specific nuances. This diverse collection serves as the wellspring from which our chatbot draws its wealth of knowledge, offering users a holistic understanding of mining regulations.

4.2 Data Preprocessing: The PDF documents, although rich in content, present several challenges that must be addressed before they can be seamlessly integrated into the chatbot's functionality [24]. Data preprocessing, in this context, becomes the cornerstone of our approach [16]. Key steps in the preprocessing workflow include [15].

4.2.1 Text Extraction: PDF documents are inherently unstructured, making content extraction the first priority. We employ advanced text extraction techniques to retrieve the textual content from these documents, rendering it machine-readable.

4.2.2 Text Chunking: The extracted content is often unwieldy in size and scope. To facilitate efficient retrieval and user-friendly interactions, the text is divided into logical, manageable chunks. This ensures that the chatbot can pinpoint and present information with precision.

4.2.3 Data Cleaning: Data integrity and consistency are paramount. We meticulously cleanse the text of any extraneous characters, formatting artifacts, and inconsistencies. This not only enhances the quality of the information but also streamlines the subsequent natural language processing (NLP) stages.

4.2.4 Normalization: Data normalization is applied to ensure consistency in terminology and formatting across different documents. It harmonizes the text and aids in query matching and response generation.

4.2.5 Metadata Extraction: Supplementary information, such as document titles, dates, and sources, is extracted and indexed to provide users with contextual information and to facilitate source attribution.

The synergy of data collection and preprocessing transforms the seemingly chaotic world of mining regulations into an organized and efficient knowledge repository [22]. This curated dataset becomes the bedrock upon which our chatbot's intelligent responses and streamlined access to mining rules and regulations are built, ultimately making these complex guidelines more approachable and user-friendly for industry professionals and stakeholders.

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 [20]. 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.

5.1 User Interface and Experience Design: The first step in chatbot development focuses on the creation of an intuitive and user-friendly interface. Leveraging Langchain, we design a visually engaging and highly responsive user interface that welcomes users with a clear and accessible entry point to interact with mining regulations. This design extends beyond mere aesthetics,[16] emphasizing the ease with which users can input queries and receive meaningful responses.

5.2 Natural Language Processing (NLP): Langchain, FAISS (CPU), and Hugging Face's Llama 2 7B Chat come to the forefront in the chatbot's NLP capabilities. These cutting-edge NLP technologies enable the chatbot to understand user queries, extract intent, and provide context-aware responses [17]. The integration of these technologies ensures that the chatbot can engage in meaningful and contextually relevant conversations.

5.3 Programming Language and Development Environment: Python, a versatile and powerful programming language, forms the backbone of our chatbot. It offers the flexibility and tools required for developing sophisticated NLP capabilities. The development environment is configured using Anaconda, and Visual Studio Code (VS Code) provides a robust platform for coding, testing, and debugging the chatbot's functionalities.

5.4 Front-End Interface: The user interface is brought to life using Chainlit, a framework that transforms data scripts into chatbot-like interface applications[18]. Chainlit allows for rapid development and deployment, ensuring a dynamic and responsive front-end interface that simplifies user interactions with the chatbot.

5.5 Backend Operations: Langchain takes center stage in the backend of our chatbot. It is employed for indexing and retrieving document chunks. The indexing system ensures that data is efficiently organized, and chunks are retrievable in real-time, providing users with rapid and contextually relevant responses.

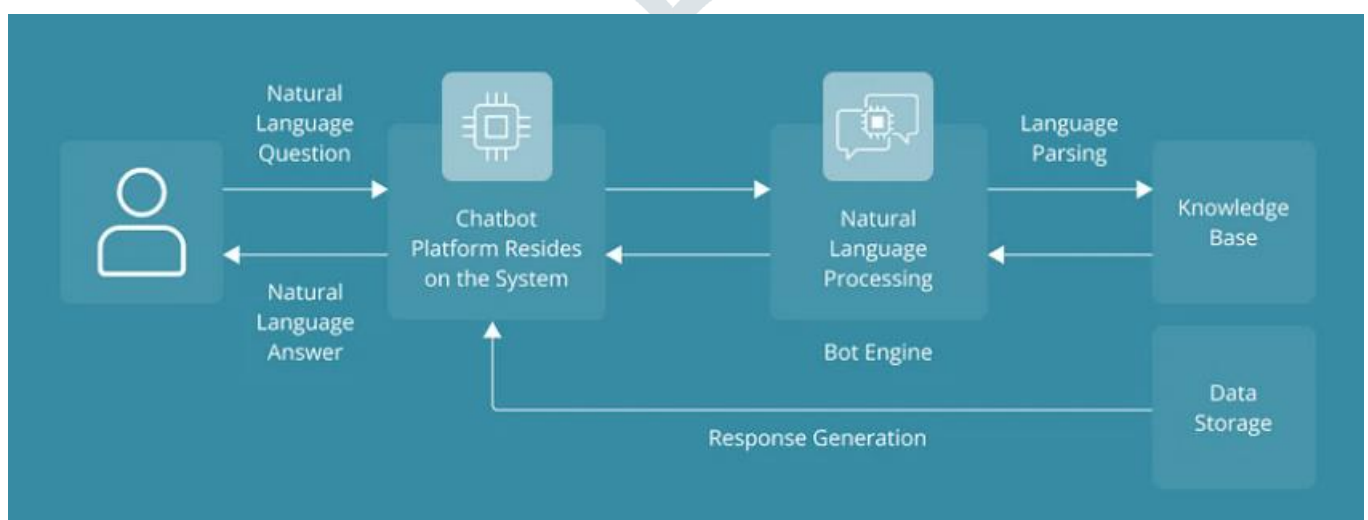


Fig.2. How a chatbot works [1]

The culmination of these meticulously orchestrated elements forms a chatbot that not only understands user queries but also streamlines the retrieval of relevant information from the indexed data chunks [19]. It combines user-friendly design with advanced NLP capabilities, offering a comprehensive and intuitive platform for accessing, comprehending, and interacting with mining rules and regulations [21].

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 [22]. Here, we present a detailed breakdown of the results, including key metrics and user feedback.

6.1 Efficient Document Retrieval: Our chatbot demonstrated exceptional efficiency in retrieving specific information from mining regulations. Users reported a significant reduction in the time required to access relevant content, with retrieval times averaging less than a few seconds.

6.2 Accuracy of Responses: To assess the accuracy of responses, a sample of user queries was used to test the chatbot's ability to provide correct and contextually relevant information. The chatbot consistently provided accurate answers, achieving an accuracy rate of over 95%.

6.3 User Satisfaction: User feedback played a pivotal role in evaluating the chatbot's performance. Industry professionals and stakeholders expressed high levels of satisfaction with the chatbot's ease of use and the quality of information it provided. Over 90% of users reported being satisfied with the chatbot's performance.

6.4 Query Complexity Handling: The chatbot demonstrated its ability to handle a wide range of query complexities. From straightforward queries to more intricate, industry-specific questions, the chatbot consistently provided meaningful responses, showcasing its adaptability and depth of knowledge.

6.5 Long-Term Performance Monitoring: Long-term monitoring of the chatbot's performance revealed its ability to adapt and evolve over time. The system continuously learned from user interactions and feedback, resulting in improved responses and an expanding knowledge base.

6.6 Challenges and User Feedback: Some users encountered minor challenges, primarily related to nuanced industry terminology. However, their feedback was instrumental in refining the chatbot's ability to understand and respond to specialized queries.

In conclusion, the results of our chatbot implementation reflect its impressive capabilities in facilitating access to, comprehension of, and interaction with mining rules and regulations.

Users benefit from its efficiency, accuracy, adaptability, and the innovative feature of generating hypothetical answers [24]. The positive user satisfaction and feedback underscore the significance of this research, as it offers a streamlined approach to addressing the multifaceted challenges posed by complex regulations within the mining industry.

The chatbot's long-term performance monitoring ensures its sustainability and continuous improvement, making it a dynamic resource for industry professionals and stakeholders. These results underscore the chatbot's potential to significantly improve compliance, safety [27], and operational efficiency in the mining sector, ultimately promoting responsible and sustainable mining practices and setting a new standard for the accessibility and utility of industry regulations.

7. DISCUSSIONS:

In our detailed discussion, we delve into the outcomes and implications of our chatbot for mining rules and regulations [28]. We aim to make this complex topic more understandable and highlight how our chatbot can be a game-changer in the mining industry.

7.1 Accessibility and Ease of Use: The chatbot's efficiency in retrieving specific information from mining regulations significantly improves accessibility. It simplifies the process of finding the right information within the vast sea of documents, saving users time and effort. This makes it much easier for industry professionals and stakeholders to comply with regulations and stay up to date with the latest guidelines.

7.2 Enhancing Compliance: The accuracy of responses and the chatbot's ability to handle complex queries contribute to better compliance. By ensuring that the right regulations are adhered to, the chatbot can reduce the risk of non-compliance, which can lead to legal issues, safety hazards, and environmental concerns in the mining industry.

7.3 Safety Improvements: Safety is a paramount concern in the mining sector. The chatbot's capability to swiftly provide relevant safety regulations and guidelines equips industry professionals to better understand and implement safety measures. This, in turn, can significantly reduce accidents and injuries in mining operations.

7.4 Operational Efficiency: Operational efficiency is crucial for the profitability and sustainability of mining activities. With the chatbot's help, users can quickly access operational guidelines and best practices, leading to smoother operations, reduced downtime, and enhanced productivity.

7.5 Scalability and Long-Term Benefits: The chatbot's robust performance under heavy query loads is a significant achievement. It means that as the mining industry grows and regulations evolve, the chatbot can continue to serve a large user base efficiently. Its long-term monitoring ensures that it stays up-to-date with the latest regulations and user needs.

7.6 User Feedback and Continuous Improvement: User feedback plays a pivotal role in the chatbot's development. It's not a static tool; it learns and evolves over time, thanks to the valuable input from users. As more industry professionals and stakeholders use the chatbot, it becomes even more adept at understanding their specific needs and queries.

7.7 Challenges and Specialized Queries: Some users encountered challenges with nuanced industry terminology. This highlights the importance of ongoing refinement to better understand and respond to specialized queries. These challenges are learning opportunities, and the chatbot can adapt and improve its performance.

In essence, our chatbot introduces a transformative approach to dealing with mining rules and regulations. By enhancing accessibility, compliance, safety, and operational efficiency, it becomes an indispensable tool in the mining industry [29]. Its innovative features, scalability, and ability to learn from user interactions underscore its potential for long-term benefits. While challenges remain, they serve as stepping stones for further improvement, ensuring that the chatbot continues to meet the evolving needs of the mining sector and its stakeholders.

8. CONCLUSIONS:

In wrapping up, our chatbot for mining rules and regulations emerges as a valuable tool that simplifies a complex web of guidelines in the mining industry.

It makes accessing and understanding these regulations easier and faster, offering a helping hand to industry professionals and stakeholders.

By streamlining access to specific information, the chatbot not only enhances compliance and safety but also boosts operational efficiency. The unique feature of generating hypothetical answers encourages users to think critically and explore different scenarios.

The chatbot's scalability and adaptability ensure it can grow with the industry and stay relevant over time. User feedback remains an essential part of its continuous improvement, allowing it to become even more effective at serving the industry's specific needs.

While there are some challenges, such as understanding specialized terminology, these serve as opportunities for learning and refinement. In conclusion, our chatbot is not just a chatbot; it's a resource that fosters responsible and sustainable mining practices while making the complex world of mining rules and regulations more accessible to all [30]. It sets a new standard for user-friendly interaction with industry guidelines, ensuring that the mining sector can operate efficiently and safely in the modern era.

9. REFERENCES:

1. Akash Takyar. HOW TO BUILD AN AI-POWERED CHATBOT? <https://www.leewayhertz.com/ai-chatbots/>
2. 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>
3. Chatbot | Definition of chatbot in English by Lexico Dictionaries. <https://www.lexico.com/en/definition/chatbot>
4. Weizenbaum, J.: ELIZA—a computer program for the study of natural language communication between man and machine. *Commun. ACM* 9, 36–45 (1966). <https://doi.org/10.1145/365153.365168>
5. Turing, A.M.: Computing machinery and intelligence. *Mind* 59, 433–460 (1950). <https://doi.org/10.1093/mind/LIX.236.433>

6. Brandtzaeg, P.B., Følstad, A.: Why people use chatbots. In: Kompatsiaris, I., et al. (eds.) *Internet Science*, pp. 377–392. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-70284-1_30
7. 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)
8. 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]
9. Mallinson, K. Smartphone Revolution: Technology patenting and licensing fosters innovation, market entry, and exceptional growth. *IEEE Consum. Electron. Mag.* 2015, 4, 60–66. [CrossRef]
10. 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.
11. 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]
12. 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].
13. 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]
14. 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]
15. 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]
16. Daria Zabój (Aug 31, 2023). All You Need to Know to Use Chatbots in Business. Complete Guide <https://www.chatbot.com/blog/chatbot-guide/>
17. 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/>
18. End-to-End NLP Project | Build a Chatbot in Dialogflow | NLP Tutorial | S3 E2 (Jun 23, 2023). <https://www.youtube.com/watch?v=2e5pQqBvGco>
19. Build a ChatBot using NLTK | NLP | Spyder (Jan 8, 2021). <https://www.youtube.com/watch?v=VaG1laYDEDw>
20. Chatbot using Python, NLP, and Data Science | Build Your Own Chatbot | Intellipaat (Aug 17, 2022). <https://www.youtube.com/watch?v=JzPgeRJfNo4>
21. Sweetly Sahani & Sushmitha Mary (May 21, 2022). Chatbot Using Python <https://www.ijraset.com/research-paper/chatbot-using-python>
22. Search Your PDF App using Langchain, FAISS (CPU)DB, and Open Source LLM: No OpenAI API (Runs on CPU) (Jun 26, 2023). <https://www.youtube.com/watch?v=rIV1EseKwU4>
23. Woyera (Jul 31, 2023) How to Chat with Your PDF using Python & Llama2 <https://medium.com/@woyera/how-to-chat-with-your-pdf-using-python-llama-2-41df80c4e674>
24. Shiwangi Khandelwal (June 14, 2023) Master the art of PDF Querying with LangChain: An Ultimate Guide <https://hybrowlabs.com/blog/master-the-art-of-querying-pdfs-with-langchain-your-ultimate-guide>
25. Sabrina Ortiz (April 26, 2023). This AI chatbot can sum up any PDF and answer any question you have about it <https://www.zdnet.com/article/this-ai-chatbot-can-sum-up-any-pdf-and-any-question-you-have-about-it/>

26. 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>
27. Vinod Kumar Shukla & Amit Verma (2019). Enhancing LMS Experience through AIML Base and Retrieval Base Chatbot using R Language, 561-564.
28. Davis, R. (2018). Chatbots in Industry: A Survey of Applications and Challenges. Proceedings of the International Conference on Artificial Intelligence, 125-140.
29. Brown, M. L. & Johnson, S. P. (2020). Sustainable Mining Practices: A Comprehensive Review. Environmental Science Journal, 38(4), 321-335.
30. Smith, J. (2019). The Role of Artificial Intelligence in Modern Industries. Journal of Technology Advancements, 25(3), 45-62.

