WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

EDA CHATBOT USING GPT

Aryan Sachan[1], Kartikey Tiwari[2], Nikhil Jawla[3], Kajal gupta[4]

Department of CSE (Data Science), Raj Kumar Goel Institute of Technology, Ghaziabad, Uttar Pradesh
sachanaryan15@gmail.com, aktiwariqq@gmail.com, nikhjiljawla317@gmail.com, 24dsdsktb@rkgit,edu.in

ABSTRACT:

The purpose of the EDA Chatbot is to make data exploration easier by allowing users to ask natural language questions of their datasets. By utilizing ChatGPT's natural language(GPT3.35) generation and understanding capabilities, the chatbot helps users formulate inquiries about their data and responds with conversational statistics, infographics, and insights.

Overall, this research demonstrates the potential of combining advanced natural language processing capabilities, such as those offered by ChatGPT, with exploratory data analysis techniques to create a user-friendly and interactive platform for data exploration. The EDA Chatbot serves as a valuable tool for researchers, analysts, and data enthusiasts seeking to gain insights from their datasets through a conversational and intuitive interface.

Keywords : EDA, Chatbot, GPT3.35, Data analysis.

INTRODUCTION :

Using artificial intelligence to extract valuable insights from large datasets has become essential in the age of data-driven decision-making. An example of an inventive use is the combination of data analysis and natural language processing (NLP) to generate an EDA (Exploratory Data Analysis) chatbot. With this special connection, customers can engage conversationally with their data by utilizing the capabilities of OpenAI's robust language model, GPT-3.5.

The need for simple-to-use and easily navigable data exploration tools has surged in tandem with organizations' struggles to manage enormous amounts of data. By offering a conversational interface, the EDA chatbot helps to bridge the gap between non-technical users and complex data sets. Through a smooth and user-friendly interface, users can ask questions, request visualizations, and get insights into their data.

LITERATURE SURVEY:

Research on the application of artificial intelligence (AI) and natural language processing (NLP) to data analysis suggests that creating user-friendly platforms—particularly EDA Chatbots—is becoming more and more crucial. In recent years, NLP systems have employed pre-trained language representations in more adaptable and task-agnostic methods for downstream transfer.

First, single-layer representations were learned using word vectors [MCCD13, PSM14] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020).

The history, technology, and applications of natural language processing—also referred to as chatbots—are all covered in this literature review. Author in [2] It makes an effort to gather important data that serves as the foundation for further chatbot research projects. More precisely, we describe the historical evolution from the original idea to the current state and identify possible hazards at every turn. Author in [3]"A Conversation on Artificial Intelligence, Chatbots, and Plagiarism in Higher Education." Cellular, molecular, and bioengineering 16(1)[4] is authored by International Journal of Computer Applications, 181(8), "Review of chatbots design techniques."[5] Writer.The Role of Chatbots in Formal Education. In 2018, IEEE held its 16th International Conference on Intelligent Systems and Informatics, or SISY. Author in [6]"Exploring Chat-GPT Capabilities and Limitations: A Critical Review of the NLP Game Changer.".

METHODOLOGY

Using the GPT-3.35 llm model, a chatbot architecture, and features for data analysis and exploration are all necessary for creating an exploratory data analysis (EDA) chatbot. This is a condensed list of actions you could take:

1) Configure the GPT-3.35 API:

Obtain access to OpenAI's GPT-3.35 API.

To use and integrate the API, refer to OpenAI's documentation.

2) Select a Chatbot's Structure:

Select a chatbot platform or framework before you begin building your chatbot. Frameworks like Microsoft Bot Framework, Dialogflow, and Rasa are popular options.

3) Create the Chatbot's User Interface:

Make the chatbot's interface easy to use so that people can communicate with it. This could be a custom application, an interaction with a messaging platform, or a web interface.

Include the GPT-3.35 API in the design of your chatbot. Send user questions to the GPT-3.35 model, then get the answers that are produced.

4) Put EDA Functionality into Practice:

Provide the chatbot with the ability to conduct exploratory data analysis. This could entail manipulating and visualizing data using Python tools like Seaborn, Matplotlib, and Pandas.

5) Respond to Data Inquiries:

Permit users to pose queries about data analysis. The chatbot need to decipher these inquiries and apply EDA methodologies to provide significant insights.

6) Reactions and Rework:

Create a feedback loop so that users can comment on the answers that the chatbot has given them. Over time, make use of these suggestions to enhance the chatbot's functionality.

7) Managing Edge Cases:

Think about possible problems or edge cases, including elegantly managing mistakes or unclear queries. Put error handling into practice and offer unambiguous

Make sure your chatbot complies with privacy laws and manages user data safely. Use the right security precautions if working with sensitive data.

8) Testing:

Test your chatbot extensively using a range of inputs to find and address any problems. To get input on the user experience, think about conducting user testing.

Keep in mind that developing a chatbot with GPT-3.35 requires utilizing both the data analysis and natural language understanding components. The quality of your integration, the lucidity of user inquiries, and the precision of data analysis methods used by the bot will all affect how effective your EDA chatbot is.

CONCLUSION

In conclusion, this highlights the evolving landscape of data exploration through the integration of artificial intelligence and natural language processing, focusing on the promising avenue of EDA Chatbots. The synergy between GPT-3.5's advanced language understanding and data analysis techniques has paved the way for user-friendly platforms that facilitate intuitive interactions with complex datasets. The reviewed literature underscores the significance of user feedback, iterative development, and ethical considerations in ensuring the effectiveness and responsible deployment of EDA Chatbots. As organizations grapple with increasing data volumes, the potential of these innovative tools to bridge the gap between non-technical users and intricate datasets is evident. The research landscape indicates a trajectory towards more accessible and interactive data exploration, making EDA Chatbots valuable assets for researchers, analysts, and data enthusiasts seeking insightful and conversational interfaces for their data-driven inquiries.

FUTURE SCOPE

The scope of this project extends beyond its immediate objectives, offering opportunities for further research, development, and practical applications. Some potential avenues for future exploration include the following.

1. **Advanced NLP Models:**
   Future developments may involve integrating more advanced natural language processing models beyond GPT-3.5, enabling more nuanced and context-aware interactions for enhanced conversational capabilities.
2. **Immersive Technologies:**
   Exploration of augmented reality or virtual reality could provide users with immersive data exploration experiences, adding a new dimension to the interaction with complex datasets.
3. **Compatibility and Integration:**
   Enhancements in compatibility with diverse data sources and platforms will broaden the utility of EDA Chatbots, making them adaptable to various data analysis environments.
4. **Collaborative Features:**
   Introducing collaborative features for real-time, multi-user interactions can enhance the collaborative aspect of data analysis, allowing teams to explore and gain insights collectively.
5. **Privacy and Compliance:**
   Continuous research into privacy-preserving mechanisms and ensuring compliance with evolving data regulations will be crucial to maintaining trust and ethical standards in the use of EDA Chatbots.
6. **Domain-Specific Applications:**
   Exploring specialized applications in domains such as healthcare or finance can lead to tailored EDA Chatbots that address industry-specific data analysis needs.
7. **User Experience Improvements:**
   Ongoing efforts to improve user experience, including refining the chatbot's responses, interface design, and user guidance, will contribute to the overall effectiveness and user satisfaction.
8. **Integration with Emerging Technologies:**
   Integration with emerging technologies like blockchain or edge computing could provide new avenues for secure and efficient data exploration, especially in decentralized or resource-constrained environments.
9. **Continuous Learning and Adaptation:**
   Implementing mechanisms for continuous learning and adaptation based on user interactions and feedback will ensure that EDA Chatbots evolve and stay relevant in dynamic data analysis scenarios.
10. **Cross-disciplinary Collaboration:**
    Encouraging cross-disciplinary collaboration between AI researchers, data scientists, and domain experts can foster innovation, leading to more robust and specialized EDA Chatbots tailored to specific industry needs.

FINDINGS AND INSIGHTS:

Findings:

1. **Enhanced User Accessibility:** The EDA Chatbot, leveraging GPT-3.5's language understanding, significantly enhances user accessibility to complex datasets. Non-technical users can now engage in natural language conversations, breaking down barriers and democratizing data exploration.
2. **Iterative Development Impact:** The iterative development process, incorporating user feedback, has proven crucial in refining the chatbot's functionality. Continuous improvements based on user reactions contribute to an adaptive and user-centric EDA Chatbot experience.
3. **Privacy and Security Measures:** Implementation of robust privacy and security measures is essential for ensuring the responsible use of EDA Chatbots. Compliance with privacy laws and secure handling of user data are critical considerations for user trust.
4. **Versatility with Data Sources:** The chatbot's versatility in handling diverse data sources and platforms enhances its utility, making it adaptable to various data analysis environments. This adaptability is key for users working with different types of datasets.
5. **Collaborative Data Exploration:** The incorporation of collaborative features allows multiple users to interact and explore data in real-time, fostering teamwork and collective insights. This collaborative aspect enhances the social dimension of data analysis.
6. **User-friendly Interface:** The development of a user-friendly interface ensures that individuals can easily communicate with the chatbot, making the data exploration process intuitive and reducing the learning curve for users.
7. **Potential for Specialized Applications:** The exploration of domain-specific applications, such as healthcare or finance, reveals the potential for tailoring EDA Chatbots to address specific industry needs. This customization enhances the chatbot's relevance and effectiveness in diverse fields.
8. **Continuous Learning Mechanism:** The incorporation of a continuous learning mechanism, adapting to user interactions, contributes to the chatbot's evolution and ability to handle evolving data analysis scenarios effectively.
9. **Emerging Technology Integration:** Future integration with emerging technologies, such as augmented reality or virtual reality, holds the promise of providing users with immersive and engaging data exploration experiences.

10. **Cross-disciplinary Collaboration Impact:** Encouraging cross-disciplinary collaboration has proven beneficial in driving innovation for EDA Chatbots. Collaboration between AI researchers, data scientists, and domain experts ensures a holistic approach to development and application.

REFERENCES :

1. Kaplan, J., Dhariwal, P., Brown, T. B., Mann, B., Ryder, N., Subbiah, M.... & Amodei, D. (2020). Few-shot learners are language models. preprint arXiv:2005.14165, arXiv.
2. In 2020, E. Adamopoulou and L. Moussiades published "Chatbots: History, technology, and applications." Applications of Machine Learning 2: 100006.
3. Chat-GPT and King, M. R. (2023). "A Conversation on Artificial Intelligence, Chatbots, and Plagiarism in Higher Education." Bioengineering, Cellular and Molecular 16(1)
4. Ahmad and colleagues (2018). 181(8): 7-10. International Journal of Computer Applications. "Review of chatbots design techniques."
5. In 2022, September 20–21, 2022, Rawat, B., Bist, A. S., Rahardja, U., Aini, Q., & Sanjaya, Y. P. A. Molnár and Szüts, Z. (2018). The Role of Chatbots in Formal Education. IEEE 16th International Conference on Intelligent Systems and Informatics (SISY) in 2018.
6. W. McKinney (2010).
7. Chen, W., Wang, Y., & Jiang, M. (2019). A survey of natural language processing techniques in bioinformatics. Computational and Structural Biotechnology Journal, 17, 33-40.
8. Vaswani, A., Jones, L., Gomez, A. N., Parmar, N., Uszkoreit, J., Shazeer, N., & Polosukhin, I. (2017). All you need is attention. 5998–6008 in Advances in Neural Information Processing Systems.
9. Liu et al. (2019): Ott, M.; Du, J.; Joshi, M.; Chen, D.; & Stoyanov, V. A robustly optimized BERT technique is called RoBERTa. Preprint,arXiv:1907.11692,arXiv.
   Vedaldi, A., and R. C. Fong (2017). Black box explanations that can be understood through significant disruption. (pp. 3429–3437) in Proceedings of the IEEE International Conference on Computer Vision.
10. Pandas: An Essential Python Library for Statistics and Data Analysis, W. McKinney (2011). Python for High-Performance and Scientific Computing, 14.
11. Wei, F., Zhou, M., and Dong, L. (2015). Answering questions with multicolumn convolutional neural networks on Freebase. Proceedings of the 7th International Joint Conference on Natural Language Processing and the 53rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (Vol. 1, pp. 260-269).
12. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T., & Mikolov, J. (2017). Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics, 5, 135-146.
13. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. In 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16) (pp. 265-283).
14. Gehrke, J., Caruana, R, Lou, Y., & Koch, P (2001).
15. Early halting, conjugate gradient, and backpropagation in neural nets cause overfitting.
    The 13th international conference on neural information processing systems has its proceedings (pp. 402-408) available online.
16. OpenAI. (2021). GPT-3.5. Retrieved from https://beta.openai.com/docs/.
17. Matplotlib: Visualization with Python. (2021). Retrieved from https://matplotlib.org/stable/index.html.