RAG CHATBOT

AIM

The aim of this project is to develop a chatbot utilizing the Retrieval-Augmented Generation (RAG) architecture for backend processing. The frontend of the chatbot will be built using Flask for web framework, along with HTML and CSS for user interface design. The project also incorporates Mistral AI model for enhancing the capabilities of the RAG architecture.

INTRODUCTION

With the increasing demand for conversational AI systems, chatbots have become integral in various domains such as customer service, healthcare, education, and more. The RAG architecture, which combines retrieval-based and generative-based approaches, offers improved performance in understanding user queries and generating contextually relevant responses. Integrating Mistral AI model further enhances the chatbot's ability to comprehend complex queries and produce more coherent responses.

OBJECTIVE

1. Develop a chatbot utilizing RAG architecture for backend processing.
2. Implement Flask, HTML, and CSS for the frontend interface.
3. Integrate Mistral AI model to enhance the chatbot's conversational capabilities.
4. Evaluate the performance of the chatbot in terms of response accuracy, coherence, and user satisfaction.

MODEL ARCHITECTURE AND IMPLEMENTATION

The chatbot architecture consists of two main components: frontend and backend.

Frontend (Flask, HTML, CSS)

1. Flask: Flask is a micro web framework for Python used to build web applications. It provides tools, libraries, and technologies for developing web applications.
2. HTML: HTML is used for structuring the content of web pages.
3. CSS: CSS is used for styling the HTML elements, providing visual aesthetics to the web pages.

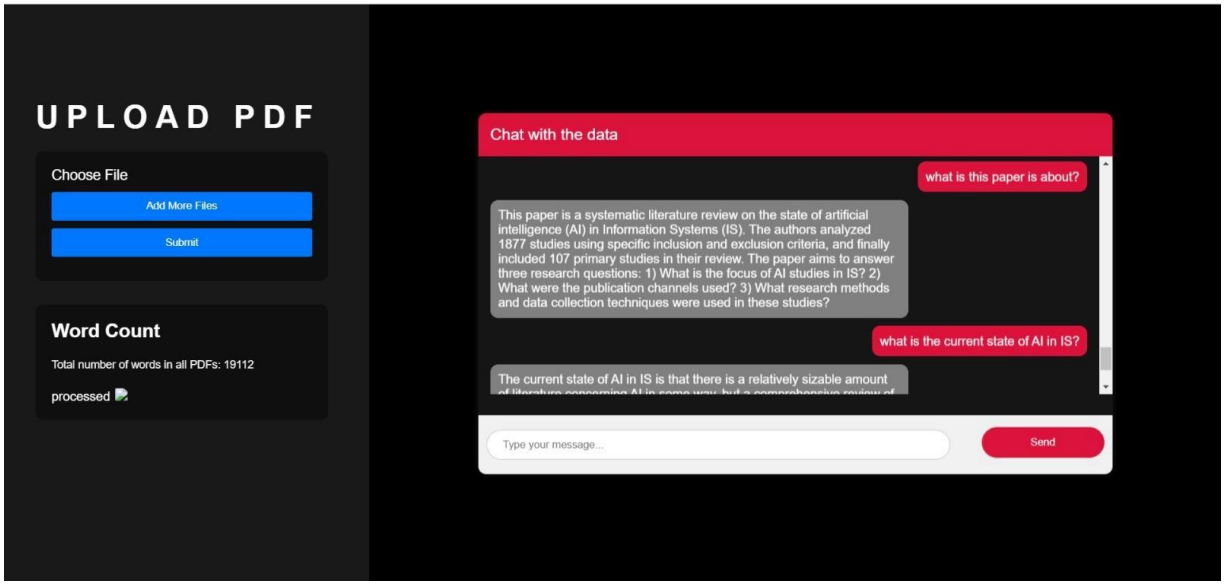
Backend (RAG with Mistral AI Model)

1. Retrieval-Augmented Generation (RAG): RAG architecture combines retrieval-based methods, where responses are retrieved from a pre-existing database, and generative-based methods, where responses are generated from scratch.
2. Mistral AI Model: Mistral AI model is integrated into the RAG architecture to enhance the chatbot's ability to understand user queries and generate coherent responses.

Evaluation

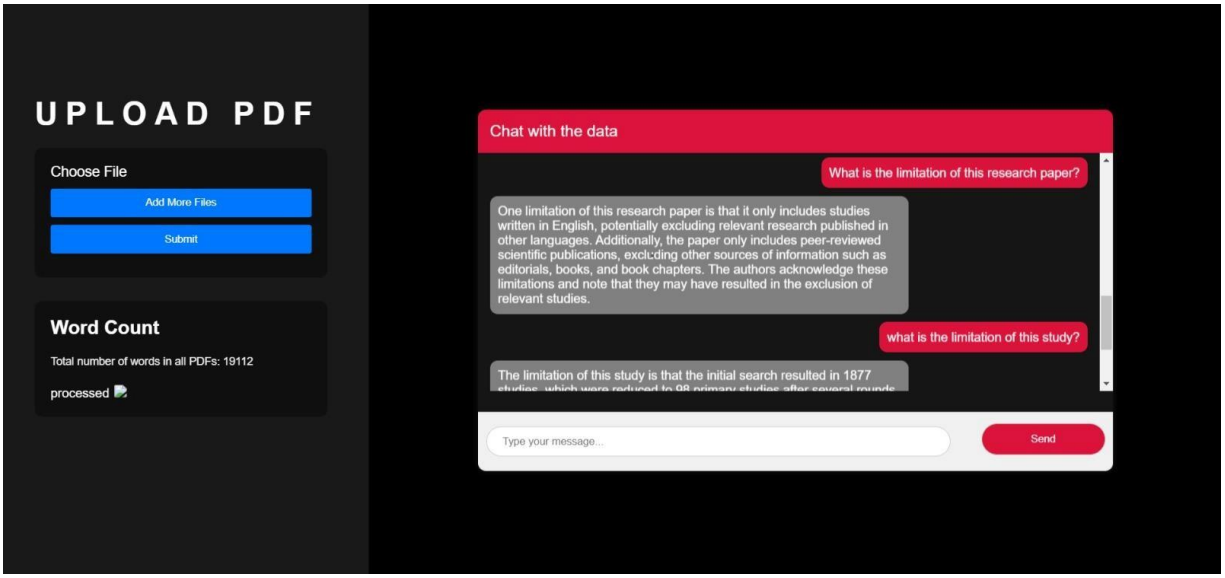
Evaluation of the chatbot is conducted based on the following metrics:

1. Response Accuracy: Percentage of queries answered correctly.
2. Coherence: Assessment of the logical flow and relevance of responses.
3. User Satisfaction: Feedback collected from users regarding their experience with the chatbot.



RESULTS

The chatbot demonstrates promising performance in terms of accuracy, coherence, and user satisfaction. The integration of the Mistral AI model significantly improves the chatbot's conversational capabilities, leading to more accurate and contextually relevant responses.



CONCLUSION

In conclusion, the project successfully develops a chatbot using RAG architecture with Flask, HTML, and CSS frontend. The integration of the Mistral AI model enhances the chatbot's performance, making it more effective in understanding user queries and generating coherent responses. Further improvements and optimizations can be explored to enhance the chatbot's capabilities and user experience.

REFERENCE

Flask - https://flask.palletsprojects.com/en/3.0.x/tutorial/templates/

- https://python-adv-web-apps.readthedocs.io/en/latest/flask_forms.html

Langchain and Rag

- https://python.langchain.com/docs/expression_language/cookbook/retrieval

- https://www.youtube.com/watch?v=LBNpyjcbv0o

(Submitted by Tiruvikraman,Surya,Anto Nikkson Dated 27-03-2024)