INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

Build a RAG System on “Leave No Context Behind” Paper

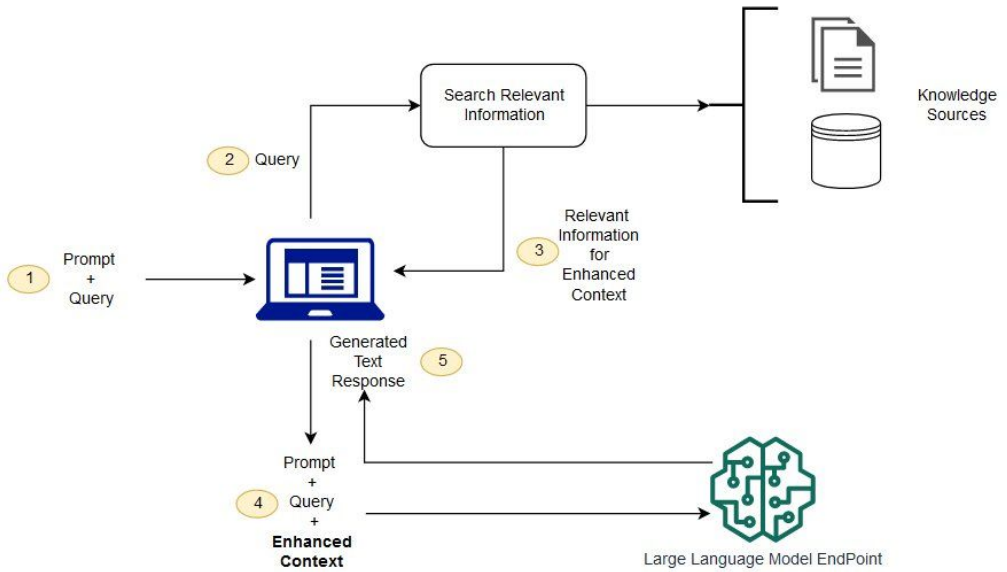
Prepared By
Lingerkar Rithikha - IN1240029

Introduction

In this report, we present the development of a Retrieval-Augmented Generation (RAG) system leveraging LangChain framework to harness the capabilities of LLMs like Gemini 1.5 Pro for answering questions related to the "Leave No Context Behind" paper published by Google.

Objective

The primary objective of this project is to create an AI-powered system capable of providing insightful answers to questions regarding the "Leave No Context Behind" paper by integrating external data sources such as PDF documents containing the paper's content.



INNOMATICS RESEARCH LABS

Methodology

1. Setting Up Environment: Begin by setting up the development environment, including installing necessary libraries and tools such as Streamlit, LangChain, and Google Generative AI.
2. Data Retrieval: Retrieve the "Leave No Context Behind" paper published by Google on 10th April 2024. This can be achieved by obtaining the PDF document containing the paper's content.
3. Data Preprocessing: Preprocess the retrieved document to extract relevant information and split it into manageable chunks suitable for processing by the RAG system.
4. Model Integration: Integrate Gemini 1.5 Pro, a powerful LLM, into the RAG system. Utilize LangChain framework to establish communication between the external data (the paper) and the LLM.
5. RAG System Development: Develop the RAG system architecture using LangChain framework. This involves defining a ChatPromptTemplate, loading the Google Generative AI model, setting up input and output parsers, and configuring the RAG chain.
6. User Interface: Implement a user-friendly interface using Streamlit to allow users to input questions related to the paper and trigger the RAG system to generate answers.
7. Testing and Validation: Test the RAG system with sample questions related to the "Leave No Context Behind" paper to ensure its functionality and accuracy. Validate the answers generated by comparing them with the paper's content.

Refer the code in the below github link.
GitHub URL : https://github.com/LingerkarRithikha/RAG_APP

Results

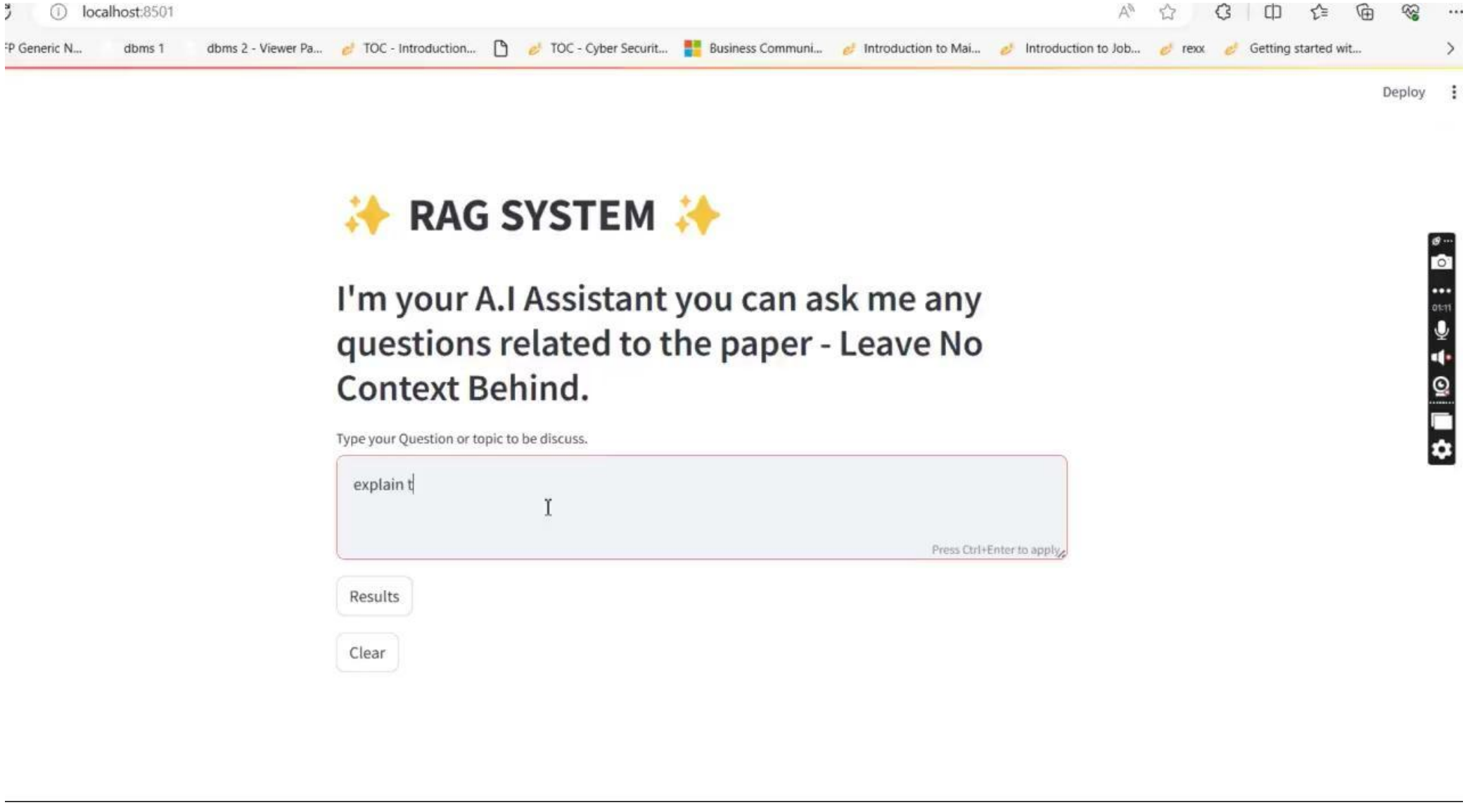
Upon completion, the RAG system successfully provides insightful answers to questions related to the "Leave No Context Behind" paper. Users can input their queries through the interface, and the system utilizes Gemini 1.5 Pro to generate responses based on the content of the paper.

Conclusion

The integration of LangChain framework with LLMs like Gemini 1.5 Pro enables the development of powerful AI systems capable of leveraging external data sources to provide contextually relevant answers. By building a RAG system for the "Leave No Context Behind" paper, we demonstrate the potential of this approach in facilitating knowledge discovery and retrieval from complex documents.

Future Work

Future enhancements to the RAG system could include refining the document retrieval process, implementing more advanced natural language processing techniques for better understanding of user queries, and expanding the system's capabilities to handle a wider range of document formats and sources.



INNOMATICS RESEARCH LABS

THANK YOU

INNOMATICS RESEARCH LABS