**CMPE-256-Hackathon**

**Title: Macro Trend Analyzer: Harnessing Language Models for Insightful Macroeconomic Analysis from User-Driven Prompts**

**Team Name:** Geek Squad

**Topic:** Copilot for achievement of UN SDGs # 1

**Team Members:** Ruchitha Reddy Koluguri(017434534) ,

Harshavardhan Raghavendra Valmiki (017442984), Suresh Ravuri(017451824)

**Git Link:** <https://github.com/ruchithareddy269/cmpe256-Hackathon/>

**Problem Statement**

Using OpenAI models, such as ChatGPT, and the LangChain library, the Macroeconomic Analysis project seeks to develop an AI-driven solution that can efficiently react to user prompts. This system is intended to retrieve data from several PDFs that users supply. Its main goals are to comprehend and evaluate user inquiries, search these documents, and produce precise, context-sensitive answers. In order to process and interpret large amounts of data in an efficient manner, ensure accurate information extraction, and generate responsive content, the challenge is to integrate sophisticated natural language processing and information retrieval techniques.

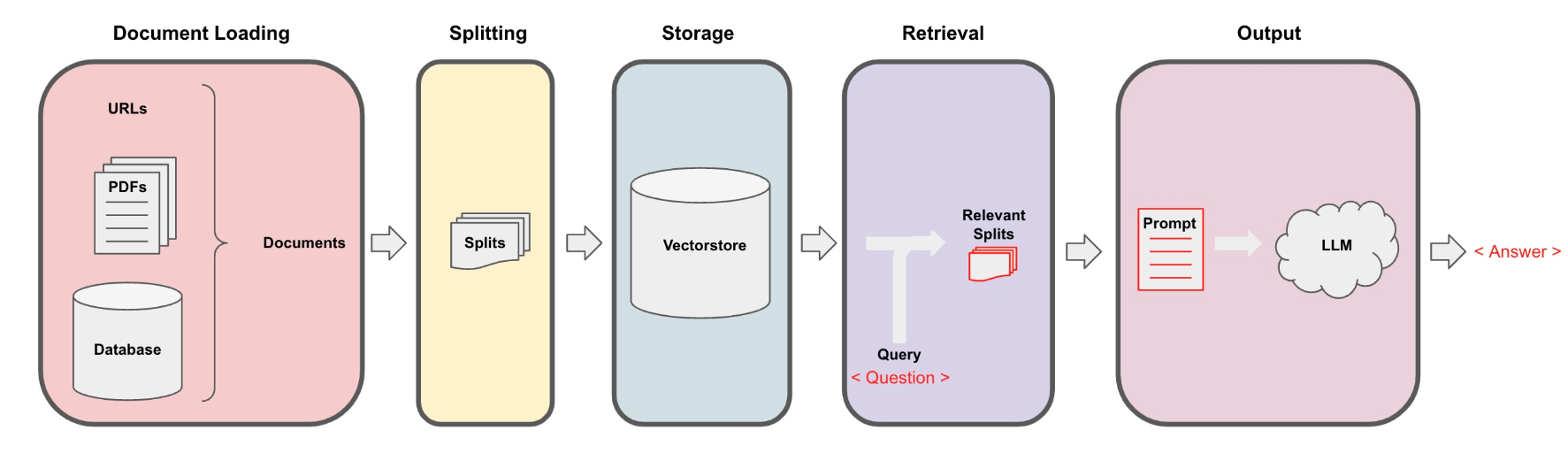
**Introduction**

This project sets out on an ambitious mission to leverage the power of cutting-edge computational models to improve our comprehension and analysis of complex datasets. Through the integration of state-of-the-art technologies like LangChain and OpenAI models like ChatGPT, the project seeks to create a novel framework that can process and comprehend large amounts of data. This project is at the forefront of using AI to inform insightful discoveries and decision-making, especially when it comes to macroeconomic analysis and sustainable development. The core of the project is its capacity to extract and handle data from various sources—mainly from various PDF documents—and display this information in a comprehensible, approachable, and intuitive way.

**Objective**

The major objective of project is to create a cutting-edge artificial intelligence tool that efficiently interprets and reacts to user inquiries by utilizing ChatGPT and other OpenAI models in conjunction with the LangChain library. The purpose of this system is to retrieve pertinent data from a number of PDF documents that users supply. One of the main goals is to develop an AI system that can analyze and interpret user queries in their entirety, search these documents thoroughly, and provide precise, contextually relevant answers. The integration of sophisticated information retrieval and natural language processing algorithms is essential to this project since it aims to handle and interpret large amounts of data in an efficient manner. This guarantees that the system is accurate in its ability to extract information and generate responses, in addition to being effective in its operational capacity. In order for the system to successfully respond to user prompts with precise and pertinent information, this balance must be reached.

**Architecture**



**What is Retrieval Augmented Generation?**

RAG is a method for adding further, frequently confidential or up-to-date facts to LLM knowledge.

Although LLMs are capable of reasoning about a wide range of subjects, their expertise is restricted to the public material available to them throughout their training period. To create AI apps that can reason over private data or data added beyond a model's cutoff date, you must add the precise information the model requires to expand its expertise. Retrieval Augmented Generation (RAG) is the process of bringing the relevant data and inserting it into the model prompt.

Retrieval Augmented Generation (RAG) is a natural language processing (NLP) method that integrates the best features of retrieval and generative based artificial intelligence (AI) models.

It contains 2 components

**1)Indexing**

**Load:** Data must first be loaded. For this, DocumentLoaders will be utilized.

**Split:** Text splitters divide lengthy documents into manageable portions. Large chunks are more difficult to search over and won't fit within a model's context window, thus this is helpful for both indexing data and feeding it into the model.

**Store:** In order to be able to search over our splits later, we need a place to index and store them. Embeddings and VectorStore are used for this.

**2)Retrieval and Generation:**

**Retrieval**: A Retriever is used to recover data from storage based on user input.

**Generation**: Using a prompt that contains the question and the retrieved data, the OpenAI generates a response.

**Technical Overview**

The integration of OpenAI and LangChain models, such as ChatGPT, which serve as the foundation of our technical framework, is central to the project. This integration is not accidental; rather, it is a deliberate decision made with the intention of maximizing the benefits of these cutting-edge technologies to improve the project's capacity for data processing and analysis.

**Data Sources and Processing**

The project focuses especially on obtaining information from PDF documents, which are difficult to process but frequently contain a wealth of useful information. The data processing pipeline, which includes steps like text extraction, normalization, and transformation to guarantee the data is ready for analysis, is painstakingly built to handle these complexities.

**Results:**

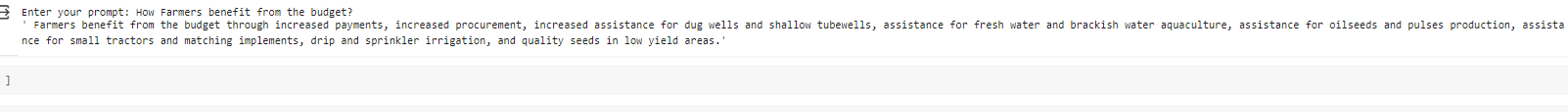
->All the prompts and responses are stored in a .**txt file** which is attached in the above git link.

**Screenshots of Input Prompts and Output Responses**

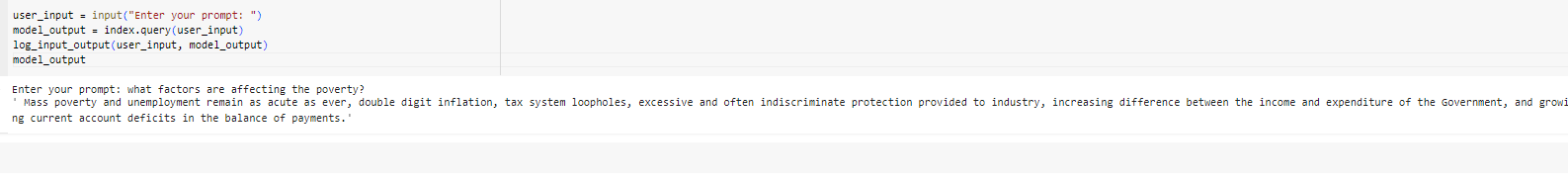
Prompt and Response 1:



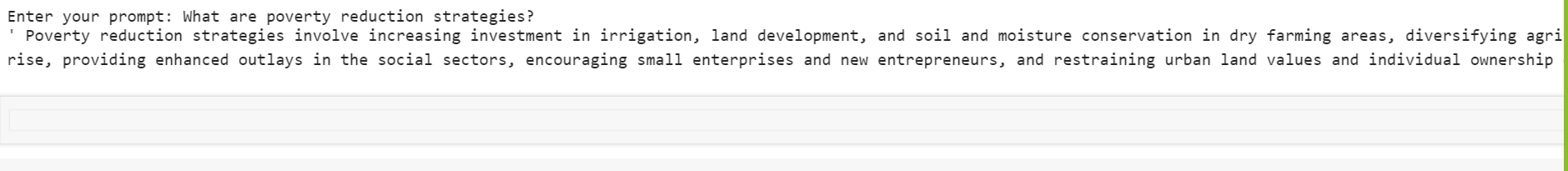
Prompt and Response 2:



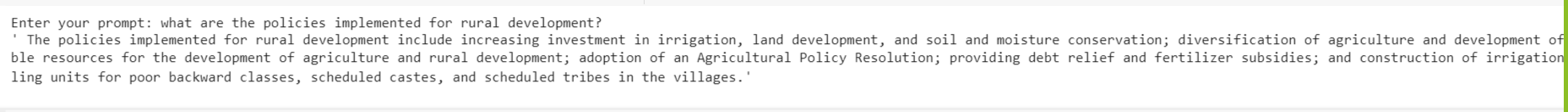
Prompt and Response 3:



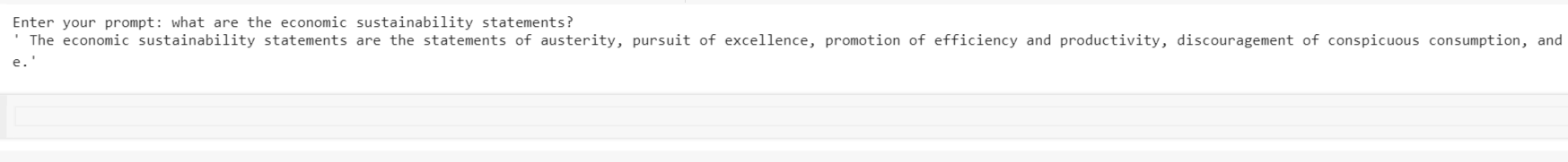
Prompt and Response 4:



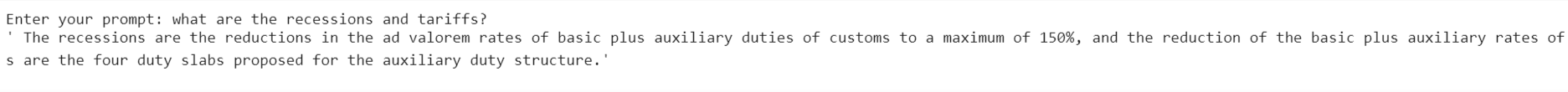
Prompt and Response 5:



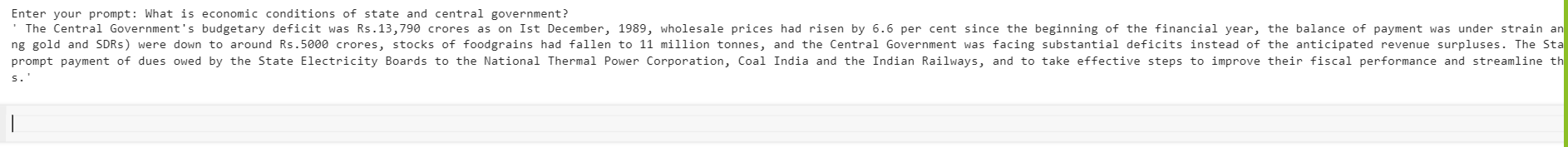
Prompt and Response 6:



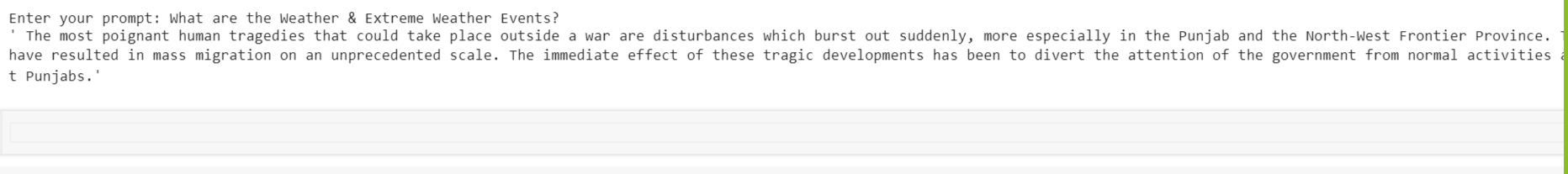
Prompt and Response 7:



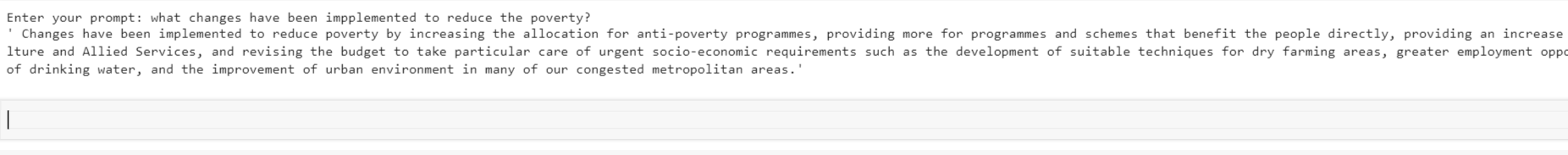
Prompt and Response 8:



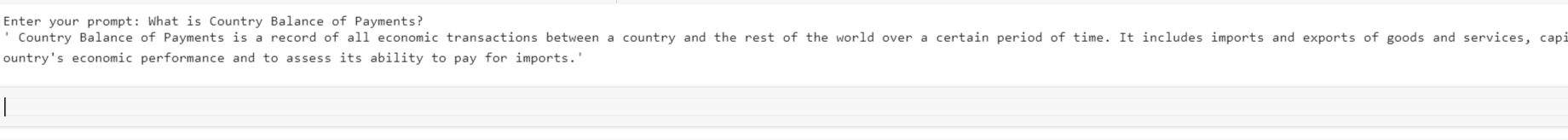
Prompt and Response 9:



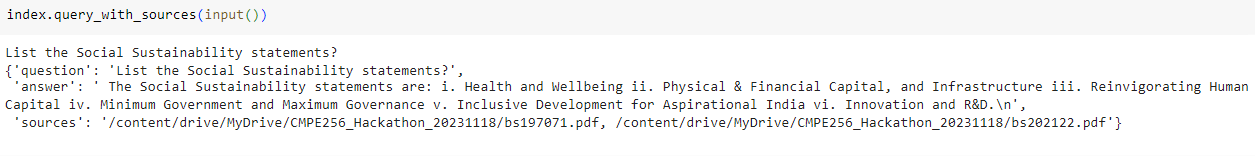
Prompt and Response 10:



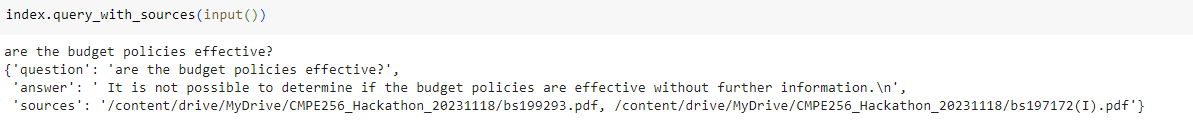
Prompt and Response 11:



**It will also print the source from where the response has been taken from by following way:**







**Conclusion:**

The project represents a critical turning point in the use of AI in macroeconomic research. It is proof of the ability of contemporary technology to convert data into useful insights, greatly advancing the fields of economic planning and sustainable development. In addition to accomplishing its goals such as providing effective answers to the provided prompts, the project creates opportunities for more study and advancement in this fascinating field.