**Project Title: Fact-Checking System Using LLM with Retrieval-Augmented Generation (RAG)**

**1. Introduction**

In the age of rapid information dissemination, fact-checking is critical. This project aims to provide users with a reliable method to verify the truth of specific information or news. By combining the capabilities of a Large Language Model (LLM) and a custom-built Retrieval-Augmented Generation (RAG) system, the system searches for evidence from trustworthy sources to generate fact-based responses. If no evidence is found, the system clearly communicates that and offers content recommendations.

The project utilizes **Streamlit** for the user interface (UI) and **Langchain** for efficient LLM integration.

**Key Features:**

* Uses **Langchain** to manage LLM tasks and responses
* Custom-built RAG system to retrieve evidence
* Sources include PDFs, Wikipedia, and reliable news websites
* Evidence-based responses, with recommendations if evidence is not available
* **Streamlit** powers the interactive and user-friendly UI

**2. Architecture and Workflow**

The system architecture consists of several key components:

* **User Interface (UI):** Built with **Streamlit**, users can enter their queries and view responses in real-time.
* **LLM (Large Language Model):** Managed using **Langchain**, this component processes the user’s query and forwards it to the RAG system for evidence retrieval.
* **RAG System (Retrieval-Augmented Generation):** The custom-built system retrieves evidence from trusted sources like PDFs, Wikipedia, and news sites.
* **Evidence Retriever:** Gathers, ranks, and filters the relevant information based on the query.
* **Recommendation System:** If no solid evidence is found, this component suggests related content or further reading.

**Workflow:**

1. A user submits a query via the **Streamlit** UI.
2. The **Langchain**-managed LLM processes the query and forwards it to the RAG system.
3. The RAG system retrieves information from trusted sources.
4. The Evidence Retriever organizes and ranks the most relevant pieces of evidence.
5. Based on the evidence found, the **Langchain** LLM generates a response.
   * If evidence exists, it provides a factual answer.
   * If no evidence is found, it returns a “no evidence” message and provides recommendations.
6. The response is sent back to the **Streamlit** UI for the user to see.

**3. RAG System Development**

The core of the project lies in the custom-built RAG system, which is responsible for retrieving reliable evidence. The system performs the following steps:

1. **Query Processing:** The query is tokenized and matched against the indexed data.
2. **Source Search:** The RAG system is connected to trusted data sources:
   * **PDF Documents:** Often government or academic publications
   * **Wikipedia:** A general source for fact-checking
   * **News Websites:** Reliable news portals providing current events
3. **Ranking & Filtering:** The Evidence Retriever ranks the results based on relevance and trustworthiness.
4. **Response Generation:** The ranked evidence is then passed to the **Langchain**-enabled LLM for generating a well-structured response.

**4. Response Logic**

The response logic is designed to ensure accuracy and transparency. It includes two main pathways:

* **If Evidence is Found:**  
  The system generates a response based on the retrieved evidence, providing sources or citations where necessary.
* **If No Evidence is Found:**  
  If no solid evidence supports the query, the system provides a “no evidence found” message and recommends related topics or content, such as similar issues or popular queries in the same domain.

**5. Challenges and Solutions**

* **Challenge: Data Reliability**
  + **Solution:** By limiting sources to trusted platforms like PDFs, Wikipedia, and established news outlets, the RAG system minimizes the risk of retrieving incorrect information.
* **Challenge: System Scalability**
  + **Solution:** The RAG system was designed to support adding more data sources in the future, allowing for better fact-checking across a wider variety of subjects.
* **Challenge: Latency in Evidence Retrieval**
  + **Solution:** Implemented efficient ranking algorithms to quickly filter and serve the most relevant data, reducing query response times.

**6. Conclusion**

This project presents a scalable, reliable fact-checking system by combining **Langchain**-managed LLM and RAG technologies. It emphasizes the importance of trustable data sources and transparency in responses, providing users with fact-based answers or clear indications when no evidence is found. Future improvements may include expanding the range of data sources and optimizing the system for broader use.