AI Developer (LLM) - Technical Task

Objective:
Build a basic RAG-based chatbot system that accepts user input (text), retrieves relevant content from a set of documents, and responds using an LLM with source references.

📝 Task Overview

You are required to build a **mini AI Query System** with the following capabilities:

🔧 Task Requirements

1. Document Ingestion

- Load and process a set of sample documents (you may use 5–10 PDFs, DOCX, or TXT files of your choice).
- Use any embedding model (OpenAI, SentenceTransformers, Cohere) to convert documents to embeddings.
- Store embeddings using a vector DB (FAISS, Chroma, or Pinecone).

2. Contextual Query Handling

- Create a basic API or script to accept a user query.
- Retrieve the top 3 most relevant document chunks using vector similarity.
- Use an LLM (OpenAI API or similar) to generate a response using the retrieved content (basic RAG loop).
- Include citations or reference source filenames in the response.

3. Tech Stack (Recommended)

- **Language**: Python
- **Frameworks**: FastAPI or Flask
- **Vector DB**: FAISS or Chroma
- **LLM**: OpenAI GPT-3.5/GPT-4 or Hugging Face models
- **Embeddings**: SentenceTransformers or OpenAI
- **Document Parsing**: PyMuPDF / PyPDF2 / python-docx

4. Bonus (Not mandatory but appreciated)

- Add basic **role-based filters** (e.g., if query is made by “Manager”, filter responses or show specific content).
- Add a **basic frontend** (can be a simple HTML/JS UI or Streamlit).
- Add a **feedback collection** mechanism to log if the response was helpful.

✅ Submission Guidelines

- Host your code in a **public or private GitHub repository** and share the link.
- Include a **README** file with:
    - Setup instructions
    - Libraries used
    - Sample queries to test
    - Notes on limitations or assumptions
- Optional: Include a short Loom video walkthrough (under 3 minutes).

🎯 Evaluation Criteria

- Code clarity and modularity
- Understanding of RAG workflow
- Ability to integrate vector DBs and LLMs
- API structuring and documentation
- Bonus: Creativity in UX or added features

📦 Deliverables

- GitHub repo link
- README.md
- Create video walkthrough and mail it to akshit.kaushik@originbluy.com