Project Title: AI Handbook Assistant using RAG

Submitted by: Nishant Roy
Course: Gen AI & LLMs

---

1. Objective

To create a simple Retrieval-Augmented Generation (RAG) system that can answer user questions based on information stored in a set of text handbooks such as “Python Basics” and “Git Commands.”

---

2. Workflow

1. Loaded 2–3 text handbooks into the system.
2. Chunked documents into smaller segments for efficient retrieval.
3. Created vector embeddings using the **SentenceTransformer (all-MiniLM-L6-v2)** model.
4. Compared user queries with all chunks using **cosine similarity** to find relevant context.
5. Simulated an LLM-generated answer grounded in the retrieved context.

---

3. Observations

Step	Outcome
Chunking	Split large documents into meaningful, searchable parts.
Retrieval	The system fetched relevant text sections for user questions accurately.
Answer Generation	The final responses were coherent and factually grounded.

---

4. Learnings

- RAG improves factual accuracy by grounding responses in provided documents.
- Embeddings and cosine similarity are powerful for text retrieval.
- Even without a real LLM, retrieval + summarization mimics real GenAI systems.

---

5. Conclusion

This project successfully demonstrates a mini-RAG pipeline that retrieves and generates grounded answers from custom handbooks, proving the power of retrieval-augmented systems in AI.