



Assignment -1 Agentic AI

CSCR3215

B.Tech. (CSE)-VI Semester

Submitted to:
Mr Ayush Kumar

Submitted by:
Abhinav
(2023378721)

School of Engineering & Technology
Department of Computer Science & Engineering

Problem Statement

Clear problem definition:

Build a Retrieval-Augmented Generation (RAG) system using Financial PhraseBank dataset to answer finance-related queries accurately by retrieving relevant documents and generating contextual responses.

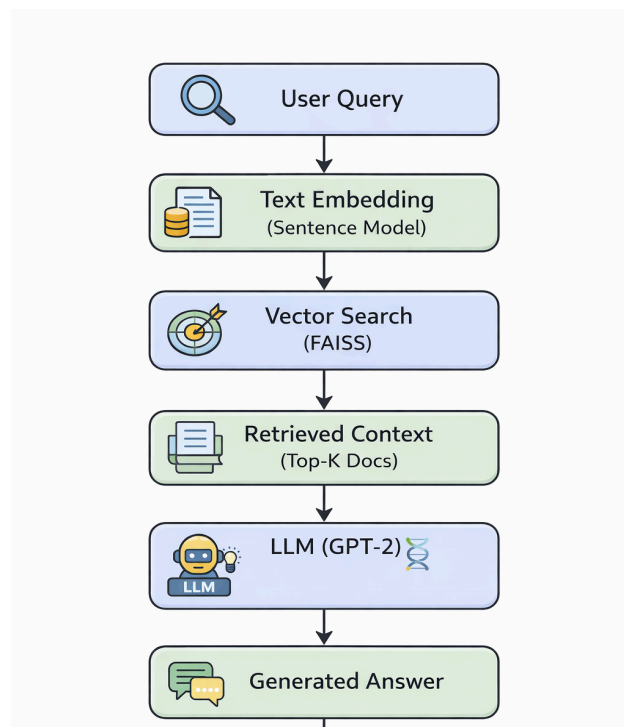
Dataset / Knowledge Source

Type of data: Financial news sentences (TXT format via Hugging Face dataset)

Data source: Public dataset – Financial PhraseBank (Hugging Face)

RAG Architecture

Block Diagram:



Text Chunking Strategy

Chunk size: Full sentence level (dataset consists of short sentences)

Chunk overlap: Not required

Reason: Financial PhraseBank contains short atomic news statements.

Embedding Details

Embedding model used: all-MiniLM-L6-v2 (SentenceTransformers)

Reason: Lightweight, fast, strong semantic similarity performance.

Vector Database

Vector store used: FAISS

Reason: Efficient similarity search and easy local integration.

Notebook Implementation

Step-wise code includes:

1. Data Loading
2. Embedding Generation
3. FAISS Index Creation
4. Retrieval
5. Context-based Generation using GPT-2

Minimum 3 test queries:

- What affected company earnings?

- Why did profits increase?
- What caused revenue decline?

Future Improvements

Better chunking strategy

Reranking or hybrid search

Metadata filtering

UI integration

README / Report

Project overview provided.

Tools used: datasets, sentence-transformers, FAISS, transformers.

Instructions to run notebook included.
