AI-Powered Research Paper Summarizer & Insight Extractor

Project Statement:
This project aims to develop a **GenAI-powered research intelligence platform** that ingests large volumes of **academic and industry research papers**, extracts key insights, and generates domain-specific summaries. By combining **RAG pipelines, semantic search, summarization models, and knowledge graph construction**, the platform helps researchers, consultants, and enterprises navigate large knowledge bases efficiently. An **interactive dashboard** will allow users to query papers, explore knowledge graphs, and export structured insights.

Outcomes:

- Automated summarization of academic and industry research.
- Extraction of key concepts, methods, and findings.
- Knowledge graph construction for relationship mapping.
- Semantic search across large research datasets.
- Dashboards for exploration and structured insight reporting.

Modules to be Implemented:

- **Document Ingestion & Parsing Layer**
  Ingests PDFs, text, and online repositories (e.g., arXiv, PubMed).
  Parses documents into structured text with metadata (title, authors, citations).

- **Summarization & Insight Extraction Engine**
  Uses **Hugging Face Transformers** (BART, T5, LLaMA) for abstractive summaries.
  Extracts domain-specific insights (key findings, methods, metrics).

- **RAG + Semantic Search Layer**
  Implements **LangChain + Pinecone/FAISS** for contextual retrieval.
  Allows natural language querying across research papers.

- **Knowledge Graph Builder**
  Extracts entities (authors, institutions, topics) and relationships.
  Stores and queries graphs using **Neo4j**.

- **Visualization & Dashboard Hub**
  React dashboards for semantic search, graph exploration, and summaries.
  Exports reports in PDF/Excel for researchers.

Milestones:

- **Milestone 1: Weeks 1–2 Data Ingestion & Parsing**
  Objective: Build ingestion pipeline for academic/industry papers.
  Tasks: Parse PDFs, normalize metadata, define schema.

- **Milestone 2: Weeks 3–4 Summarization & Insight Extraction**
  Objective: Automate summarization and concept extraction.
  Tasks: Apply Hugging Face models, validate summaries with domain experts.

- **Milestone 3: Weeks 5–6 RAG & Knowledge Graphs**
  Objective: Enable contextual search and graph building.
  Tasks: Integrate LangChain + Pinecone, build Neo4j graph of entities/relations.

- **Milestone 4: Weeks 7–8 Dashboards & Deployment**
  Objective: Deliver visual insights and deploy system.
  Tasks: Build React dashboards, enable semantic search UI, deploy platform.

Evaluation Criteria:

- **Milestone 1 Evaluation (Week 2):** Papers parsed; metadata and schema validated.

- **Milestone 2 Evaluation (Week 4):** Summaries and insights generated accurately.

- **Milestone 3 Evaluation (Week 6):** RAG + knowledge graph operational.

- **Milestone 4 Evaluation (Week 8):** Dashboards deployed; system live for research use.