Visual Document Analysis RAG

Problem Statement

Build a RAG system that can process PDFs, images, and scanned documents to extract and retrieve information from tables, charts, and mixed text-image content.

Key Requirements

- Multi-format document processing (PDF, images, scanned documents)
- Table and chart data extraction
- Mixed text-image content understanding
- OCR integration for scanned documents
- Visual element recognition and indexing

Technical Challenges

- OCR accuracy for various document qualities
- Table structure recognition and data extraction
- Chart and graph interpretation
- Layout analysis for mixed content
- Image-text correlation and context preservation

Deliverables

A fully working deployed demo (e.g., via Streamlit, Gradio, or HuggingFace Spaces)

A well-structured GitHub repository with clean code, documentation, and a README.md explaining the system

A public link to the working application

Project Scope & Guidelines

Each RAG project will focus on a specific domain such as law, healthcare, finance, education, or multimodal data processing (text, image, audio, video).

Students must:

- Use appropriate embedding models (e.g., OpenAI, HuggingFace Sentence Transformers)
- Implement retrieval using vector databases like **Chroma**, **Pinecone**, or **Weaviate**
- Design effective **chunking strategies** tailored to the data type
- Provide meaningful **retrieval-based responses** using context-aware generation
- Ensure their system has **clear UX**, **logical data flow**, and **relevance scoring**
- Evaluate with basic metrics (e.g., retrieval accuracy, latency, or RAGAS)

Submission Requirements

- GitHub repo link
- Deployed app link
- **Deadline: 3 days from the assigned day**