Intelligent PDF Summarizer & QA Bot – Explanation

◆ Use Case

- Reading long research papers and reports is time-consuming.

- Users need quick summaries and answers without reading the full document.

◆ Solution

- An **AI-powered PDF Bot** that can:

  - Extract text from PDFs (with OCR for scanned pages).

  - Create **structured summaries** with key insights.

  - Answer user queries with **page-level citations**.

◆ Workflow

1. **PDF Upload**

2. **Text Extraction** (OCR fallback if needed)

3. **Chunking & Embeddings**

4. **ChromaDB (Vector Store)**

5. **Retriever + LLaMA3 (Ollama Model)**

6. **Outputs**: Structured Summary + Q/A with Page Citations

◆ Impact

- Saves **time and effort** for students, researchers, and professionals.

- Provides **reliable answers with references** → improves trust.

- Makes research and learning **faster and more accessible**.

◆ Tools & Tech

- **LangChain, ChromaDB, Hugging Face Embeddings**

- **Ollama (LLaMA3:8B Instruct Model)**

- **PyMuPDF, pytesseract, pdf2image**

- **Streamlit for UI**