### **Architecture**

End-to-end pipeline for generating questions from PDFs

## **PDF Upload**

Select file, validate type/size • Choose page range ( $\leq$  30)



### Chunking

Recursive/Character splitter • Configurable size & overlap



#### **Text Extraction**

PyPDF per-page text • Skip empty pages



#### **Context Prep**

Normalize whitespace • Concatenate ordered chunks

#### **Embeddings**

Sentence-Transformers all-mpnet-base-v2 (768-d)



## **FAISS Index**

IP similarity • L2-normalized vectors • Metadata (pickle)

## Topics (LLM)

Concise list from indexed text • Retry + rate limit



### **Question Generation**

Mistral prompts • Types (MCQ/Short/Desc) • Difficulty

### Results

Render by topic/type • Save selected items



## **Exports**

CSV export • Styled PDF via ReportLab

# **Technology Stack**

Flask

<mark>leb framework</mark>

Mistral Al

PhiData
Agentic flow

FAISS
Vector store

Transformers

MPNet

PyPDF Parsing ReportLab

PDF export