**1.What method or library did you use to extract the text, and why? Did you face any formatting challenges with the PDF content?**

**I used PyPDFLoader from langchain\_community.document\_loaders to extract text from PDFs as it preserves document structure. Some PDFs had inconsistent formatting, especially around tables and headers, which needed cleanup.**

**2.What chunking strategy did you choose (e.g. paragraph-based, sentence-based, character limit)? Why do you think it works well for semantic retrieval?**

**I used RecursiveCharacterTextSplitter with a fixed character limit. This ensures context-rich, overlapping chunks suitable for semantic retrieval while avoiding loss of meaning between boundaries.**

**3.What embedding model did you use? Why did you choose it? How does it capture the meaning of the text?**

**I chose sentence-transformers/all-MiniLM-L6-v2 for its balance of speed and performance. It produces dense embeddings that effectively capture sentence-level semantics.**

**4. How are you comparing the query with your stored chunks? Why did you choose this similarity method and storage setup?**

**I used cosine similarity via Qdrant (a vector DB) to compare query and chunks. Qdrant offers efficient indexing with AWS hosting and retrieval for similarity-based search in production setups.**

**5. How do you ensure that the question and the document chunks are compared meaningfully? What would happen if the query is vague or missing context?  
The retriever uses semantic similarity to match chunks with the query. If the query is vague, results may be less relevant; refining the prompt or using a hybrid search could help.**

**6. Result Relevance & Improvements:  
Results are generally relevant, but quality improves with more accurate chunking, better embeddings, or a larger, more diverse document base for context coverage. For bangla it complicates things.Need more efficient bigger models to for proper bangla conversation.**

**\*\*USED QWEN3\_ 0.6B version as LLM and QUDRANT as vector database for industry level quality assurance**