**High-Level Architecture Diagram**

Here’s a textual description for a typical architecture:

1. **Frontend (User Interface):**
   * Built using **Streamlit** to provide an intuitive UI for uploading documents, querying, and displaying results.
2. **Backend Processing:**
   * **Preprocessing Module (src.preprocessing):** Extracts and preprocesses text from uploaded PDFs.
   * **Embeddings Module (src.embeddings):** Generates embeddings for text chunks using a pre-trained Sentence Transformer model.
   * **Vector Database Module (src.vector\_db):** Handles storage and retrieval of embeddings via FAISS.
3. **Persistent Storage:**
   * **Document Storage:** Uploaded PDFs are saved locally in the documents/ directory.
   * **Vector Database Storage:** Serialized embeddings are saved in a pickle file (data/vector\_db.pkl).
4. **Inference:**
   * Query embeddings are matched with stored embeddings using FAISS to fetch top results.