**Project Overview: AI Based Chatbot with Semantic Search and Citations**

**Project Overview**:  
The **Conversational QA System** integrates **Streamlit** for the web interface, **Hugging Face transformers** for text generation, **FAISS** for efficient search, and **PyPDF2** for PDF text extraction. The system allows the user to ask questions either related to general knowledge or content from an uploaded PDF document.

**Core Components**:

1. **Streamlit**:
   * Provides an interactive front-end for users to interact with the model.
   * Displays conversation history, citations, and the session title in the sidebar.
2. **Text Processing**:
   * **PyPDF2** is used to extract text from PDF documents.
   * **Sentence-Transformer** (paraphrase-MiniLM-L6-v2) is used to create embeddings for each sentence in the document.
3. **FAISS**:
   * A fast similarity search library to index sentence embeddings and retrieve the most relevant sentences based on a query.
4. **Model**:
   * **Flan-T5-Base** is used for answering questions, both for general QA and PDF-based QA (RAG system).

**How it works**:

1. **Normal QA Mode**:
   * When a user asks a question, it’s passed to a pre-trained model, which generates a detailed response.
2. **PDF-based QA Mode**:
   * After uploading a PDF, its text is processed, split into sentences, and each sentence is embedded.
   * **FAISS** indexes the embeddings, and when a user asks a question, the model retrieves the most relevant sentences to generate the answer.
3. **Sidebar**:
   * Displays the conversation history and citations for the answers (for PDF-based QA), including the page number and sentence from the document.