**Assignment - RAG-Based Chat Application**

This is an AI-powered **Retrieval-Augmented Generation (RAG)** chat application that allows users to **upload their own files** and **chat with the content**. The system creates a **FAISS index** from the uploaded document and uses a **large language model (LLM)** to answer questions based on the file’s content, enabling highly accurate and contextual conversations.

**Features -**

* 📁 **Custom File Upload**: Upload your own .pdf, .txt, or .docx files
* 🔍 **FAISS Indexing**: Automatically creates a vector store from uploaded document content
* 🤖 **Chat Interface**: Ask questions and get answers grounded in your file using natural language
* 📚 **Contextual Retrieval**: Combines document retrieval with LLM generation for accurate responses
* ⚡ **Powered by Zephyr-7B** via Hugging Face
* 🌐 **Streamlit Frontend**: Simple and interactive user interface

**Technologies Used -**

* **Python** – Backend logic and file processing
* **LangChain** – Manages RAG pipeline and prompt orchestration
* **HuggingFace Transformers** – Hosts the Zephyr-7B LLM
* **FAISS** – Efficient similarity search over document embeddings
* **Sentence Transformers (MiniLM)** – For embedding document chunks
* **Streamlit** – Interactive web app for uploading files and chatting

[⚡ Built fast *with* AI, *thanks* to AI… because how else do you impress *Build Fast with AI*? 😄]