Project Submission Confirmation

Project Title: Development of Retrieval-Augmented Generation (RAG) Model for QA Bot

Date: [13-10-2024]

Project Overview: This project involved the successful development of a Retrieval-Augmented Generation (RAG) model for a Question Answering (QA) bot. The model integrates a vector database for document embedding retrieval and a generative model to provide contextually relevant answers.

Tasks Completed: 1. RAG Model Implementation: - A RAG-based model was built to handle questions related to documents. - Initially, Pinecone was considered for document embedding storage and retrieval, but due to issues with creating an index and managing environment keys, the FAISS vector database (an open-source alternative) was used instead. - The model was tested with several queries, successfully retrieving and generating accurate answers. 2. Interactive QA Bot Interface: - Developed an interactive frontend using Streamlit that allows users to upload PDF documents and ask questions based on the document content. - Integrated backend for real-time document processing and response generation. - Implemented efficient query handling and display of retrieved document segments alongside generated answers. 3. Documentation: - Comprehensive documentation was provided detailing the model architecture, retrieval approach, and generative response methodology.

- Example interactions were included to demonstrate the bot's capabilities.

Additional Work: In addition to the initial setup, the FAISS vector database was fully integrated to enhance the efficiency and flexibility of document embedding retrieval, yielding high performance and accuracy during testing.

Deliverables:

- Colab notebook demonstrating the pipeline from data loading to question answering.
- Documentation explaining the system's architecture, setup, and usage.
- Fully functional QA bot deployed with a user-friendly interface allowing document upload and interaction.
- Source code and deployment instructions shared via GitHub.

This confirms the successful submission of the project, with the use of FAISS for vector storage.