

Deep Learning Model Development and Comparative Evaluation Report

Deepak Kumar Ayyasamy

University of Missouri – Kansas City

COMP_SCI5588: Data Science Capstone

Yugyung Lee

April 29, 2024

Deep Learning Model Development and Comparative Evaluation

Model:

1. **LLM:** Quantized Mistral-7B is used for question answering or text generation tasks.
2. **Embedding Model:** The embedding model is loaded using the **HuggingFaceInstructEmbeddings** class and the specific embedding model used is named "hkunlp/instructor-large". This embedding model is crucial for converting text inputs into numerical representations that the Mistral model can understand and process effectively.
3. **Retrieval-QA Pipeline:** The **retrieval_qa_pipeline** function sets up a question-answering pipeline. It loads the Mistral model, retrieves embeddings, and configures a retrieval-based question-answering system. This system uses a vector store (Chroma) for efficient retrieval of relevant documents and responses.
4. **Main Function:** The **main** function serves as the entry point for running the question-answering system. It accepts parameters such as the device type, whether to show source documents, whether to use history, and the model type. It initializes the QA pipeline and continuously prompts the user for input queries, providing answers based on the Mistral model's responses.

Datasets:

Source documents for training the model is Course Curriculum of UMKC.

Evaluation Results:

Aspect	UniBuddy	Gemini
Relevance to Question	10	10
Coverage of resources and support	8	10
Clarity and Coherence	10	10
Additional Information	8	10
Overall Score	9	10