Retrieval Augmented Generation(RAG) in LLM

Section 1: Theoretical Knowledge

1.1. Basics of Retrieval-Augmented Generation (RAG)
Question: Explain the concept of Retrieval-Augmented Generation (RAG). How does it differ from standard text generation methods in NLP?

1.2. Information Retrieval Techniques
Question:What are Dense Passage Retrieval (DPR) and Sparse Retrieval? Compare their strengths and weaknesses.

1.3. Transformer Models in NLP (10 Points)
Question: Briefly describe the architecture of Transformer models and their role in LLMs. How do they contribute to RAG?

Section 2: Practical Skills

2.1. Python and NLP Libraries
Task: Write a Python script that tokenizes a given text using the Hugging Face Transformers library. Use BERT's tokenizer for this task.

2.2. Basic Retrieval Task
Task: Implement a simple keyword-based retrieval system using TF-IDF. Given a set of documents and a query, return the top 3 most relevant documents.

2.3. Basic LLM Task
Generate a short paragraph on the topic "The importance of retrieval in NLP" using a pre-trained GPT model from Hugging Face.

Deadline: **18/08/2024**