

Performance Analysis of LLMs on Multi-Document Summarization

Objective: The primary objective of this assignment is to evaluate the performance of open-source Large Language Models (LLMs) on multi-document summarization tasks. The performance of LLMs will be benchmarked against a fine-tuned pre-GPT-3 model such as BART or T5.

Datasets: For this analysis, you can select any dataset relevant to multi-document summarization from the following source: Papers with Code - Multi-Document Summarization Datasets.

Tasks:

- **Dataset Selection and Pre-processing:**
 - Choose a suitable dataset from the mentioned source.
 - Pre-process and prepare the dataset for training and evaluation.
- **Evaluation of LLMs:**
 - Select an open-source LLM for the task.
 - Evaluate the LLM on the test set of your chosen dataset.
 - Document the results, including but not limited to ROUGE score, BLEU score, etc.
- **Comparison with Pre-GPT-3 Models:**
 - Fine-tune a pre-GPT-3 model, such as BART or T5, for the multi-document summarization task.
 - Evaluate the fine-tuned model on the same test set.
 - Compare and contrast the performance of the LLM and the fine-tuned pre-GPT-3 model.
 - Highlight the strengths and weaknesses of each approach.