**Assignment Number 3**

**CS-4063 Natural Language Processing**

**Fall 2024**

**Instructions:**

* Submission will be **Online** with **Screenshots** of running code
* Late submissions are not allowed.
* Solve all the questions in the given order.
* Write the programs with clarity and add comments where necessary.
* Each question carries 10 marks.
* The evaluation will be viva based with binary marking.

**Title: Text Summarization Using Large Language Models (LLMs)**

**Objective:**  
The goal of this assignment is to develop and fine-tune various Large Language Models (LLMs) for the task of text summarization. You will work with a dataset to fine-tune BERT, GPT-2, and Llama models, evaluate their performance using appropriate metrics, and compare their results.

**Public Dataset:**  
Use the following text summarization dataset. https://www.kaggle.com/code/lusfernandotorres/text-summarization-with-large-language-models/input

### Requirements:

#### 1. Fine-tuning the Models for Text Summarization:

You are required to fine-tune the following pre-trained models for the task of text summarization:

* BERT (Bidirectional Encoder Representations from Transformers): Fine-tune BERT for extractive summarization, where the model identifies the most relevant sentences from the text and uses them to create the summary.
* GPT-2 (Generative Pre-trained Transformer 2): Fine-tune GPT-2 for abstractive summarization, where the model generates a coherent summary based on the content of the input text.
* Llama (Large Language Model Meta): Fine-tune Llama for either extractive or abstractive summarization, depending on your approach.

For each model:

* Implement the fine-tuning process using your dataset.
* Use an appropriate framework like Hugging Face's Transformers or any other framework you are comfortable with.

#### 2. Evaluation Metrics:

After fine-tuning the models, evaluate their performance based on the following:

* Loss: Use the loss function (Choose it wisely and give explanation why you used this loss ) during training and report the final training loss for each model.

Additionally, you may also report on other evaluation metrics, such as ROUGE scores, if applicable.

#### 3. Sample Text and Predicted Summaries:

After training each model, take a sample text from the dataset and input it into the trained models. For each model, you should:

* Provide the original text.
* Show the predicted summary generated by each model (BERT, GPT-2, and Llama).

This will help in visually comparing the summaries produced by each model and assess their effectiveness.

#### 4. Comparison of Results:

* Report the results of each model in terms of:
  + Loss: Present the loss values during training and final evaluation for each model.
  + Summary quality: Subjectively evaluate the quality of summaries generated by each model. Consider aspects like conciseness, relevance, and coherence.
* Comparison of Summaries:
  + Compare the generated summaries across all models. Highlight any significant differences in the quality, style, or length of the summaries.
  + If possible, use metrics like ROUGE-N, ROUGE-L, or BLEU to objectively compare the summarization quality of each model.