**Assignment 1 – Problem Statement F**

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

**Instructions**

1. **Follow the instructions in each question carefully.**
2. **Only two files should be uploaded in canvas without zipping them. One is ipynb file and other one html output of the ipynb file. No other files should be uploaded**
3. **Any assignment submitted using other python IDEs are not considered for grading.**
4. **Incorrect Assignment Set submitted will not be considered.**

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

**Data set Link:**

[**https://huggingface.co/datasets/nvidia/HelpSteer**](https://huggingface.co/datasets/nvidia/HelpSteer)

**Problem Statement:**

Develop and implement a system utilizing parameter-efficient fine-tuning (PEFT) techniques to build an improved system showing the benefits of PEFT over full-model fine-tuning. The objective is to demonstrate the effectiveness of PEFT to create a highly efficient and accurate systems. The project will culminate in a detailed implementation, results, and analysis.

**Tasks:**

1. Perform necessary data preprocessing techniques. [1 Mark]
2. Explain and implement how LLM (2B to 8B parameters) over simple LM can help improve the system accuracy. [1 Mark]
3. Explain and implement how specific tokenizer and model can work well with the mentioned dataset. [1 Mark]
4. Explain and implement how sentiment analysis can be used to analyze user questions or contexts. [1 Mark]
5. Implement fine-tuning for an open-source model to improve the results. [3 Marks]
6. Show 5-10 working examples that show improvements of the accuracy of the fine-tuned model. [3 Marks]
7. Apply and fine-tune the generator model using Parameter-Efficient Fine-Tuning (PEFT). [3 Marks]
8. Implement a function to evaluate the model's performance using metrics such as accuracy and F1 score. [1 Mark]
9. Analyze the results to discuss the impact of PEFT on model performance and efficiency. [1 Mark]