Implementation Report

Implementation & Training

Each of these models was trained for 5 epochs using Adam optimizer with learning rate 5×10^{-4}.

The batch size for the training was set to 10, to avoid causing memory issues on Kaggle. The maximum length was set to 256 as well, in the GPT-2 tokenizer.

The models were trained on entire dataset.

Results of Training the Models:

1. LoRA model

Hyperparameters chosen: $\alpha = 16$ (scaling factor), $r = 8$ (rank of matrix)

		Before Training	After Training
Training Set	Avg. Loss	12.21	3.93
	ROUGE-1	29.65	42.72
	ROUGE-2	11.69	19.37
	ROUGE-L	19.30	24.56
Evaluation Set	Avg. Loss	12.46	4.31
	ROUGE-1	30.12	45.73
	ROUGE-2	12.23	22.41
	ROUGE-L	20.91	27.76

2. Traditional Fine Tuning

		Before Training	After Training
Training Set	Avg. Loss	12.21	4.68
	ROUGE-1	29.65	40.68
	ROUGE-2	11.69	18.14
	ROUGE-L	19.30	25.71
Evaluation Set	Avg. Loss	12.46	5.34
	ROUGE-1	30.12	42.33
	ROUGE-2	12.23	19.95
	ROUGE-L	20.91	27.25

3. Soft Prompt Tuning

Hyperparameters chosen: Prompt length = 10, random initialization

		Before Training	After Training
Training Set	Avg. Loss	12.21	2.87
	ROUGE-1	29.65	12.77
	ROUGE-2	11.69	5.23
	ROUGE-L	19.30	8.35
Evaluation Set	Avg. Loss	12.46	3.21
	ROUGE-1	30.12	13.58
	ROUGE-2	12.23	6.21
	ROUGE-L	20.91	10.49

Resource Utilization for each of the models:

	No of parameters		Time taken for training
	Total	Trainable	(per epoch)
Traditional Fine Tuning	124439808	38597376	26 mins 7 secs
LoRA	124882176	442368	27 mins 32 secs
Soft Prompt	124447488	7680	25 mins 56 secs

Analysis:

- **LoRA:** LoRA demonstrated the best performance in terms of ROUGE scores, particularly on the evaluation set, suggesting superior generalization and adaptation to the task. It outperformed the Traditional Fine tuning by a good margin.
- **Traditional Fine tuning:** The traditional fine tuning is competitive and works. It had lower training time than LoRA but lower performance than LoRA as well.
- **Soft Prompt:** It performed significantly bad compared to LoRA and Traditional but if trained for more epochs, it could possibly attain a competitive level as them. It had extremely few parameters, but this compromised the adaptation very much.

Best Performance: LoRA