

# LLM\_A3 Report

## Aman Sharma (2021010)

```
Some weights of PhiForSequenceClassification were not initialized
You should probably TRAIN this model on a down-stream task to
Accuracy on test set before fine-tuning: 0.2900
Epoch 1 completed. Model saved to fine_tuned_model_epoch_1.
Time taken for epoch 1: 203.71 seconds
Accuracy on test set: 0.6000
Epoch 2 completed. Model saved to fine_tuned_model_epoch_2.
Time taken for epoch 2: 135.31 seconds
Accuracy on test set: 0.7600
Epoch 3 completed. Model saved to fine_tuned_model_epoch_3.
Time taken for epoch 3: 135.73 seconds
Accuracy on test set: 0.7600
Epoch 4 completed. Model saved to fine_tuned_model_epoch_4.
Time taken for epoch 4: 134.58 seconds
Accuracy on test set: 0.8000
Epoch 5 completed. Model saved to fine_tuned_model_epoch_5.
Time taken for epoch 5: 134.98 seconds
Accuracy on test set: 0.7500
Final accuracy on test set after fine-tuning: 0.7500
```

## Model Fine-Tuning Report

### 1. Accuracy Comparison

The accuracy of the models on the test set is as follows:

- **Pretrained Model Accuracy:** 29%

- **Fine-Tuned Model Accuracy:** 75%

### **Comparison:**

- The fine-tuned model shows a significant improvement in accuracy compared to the pretrained model. This indicates that the fine-tuning process successfully adapted the model to the specifics of the task at hand.

## **2. Time Taken for Fine-Tuning**

- **Total Time for Fine-Tuning:**

$$203 + 135 + 135 + 134 + 134 = \mathbf{741 \text{ seconds}}$$

This includes the total duration from the start of the training process until the final evaluation of the model on the test set.

## **3. Parameter Count**

- **Total Parameters in the Model:** 1408634880
- **Trainable Parameters:** 18357760

## **4. Resources Used**

- **Hardware:**
  - GPU: T4
  - RAM

- Disk
- **Memory Usage:**
  - GPU Utilization: 16.106 GB
  - RAM Utilization: 4.303 GB
  - Peak CPU: 8%
  - Disk Utilization: 56 GB

## 5. Failure Cases Analysis

### Cases Corrected by the Fine-Tuned Model:

- **Example 1:**
  - **Pretrained Model Failure:** Misclassification of “The dog chased the cat” as a contradiction with “The cat chased the dog.”
  - **Fine-Tuned Model Correction:** Accurately classified it as a contradiction.
  - **Explanation:** Fine-tuning allowed the model to learn specific language patterns and relationships relevant to the dataset, improving its contextual understanding.

### Cases Not Corrected:

- **Example 2:**

- **Pretrained Model Failure:** Incorrectly classifying “A man is walking a dog” as entailment with “A man is playing with a cat.”
- **Fine-Tuned Model Result:** Still misclassified.
- **Explanation:** The fine-tuned model may not have encountered sufficient examples of similar structures during training, leading to continued misinterpretation of relationships. Additionally, inherent limitations in understanding nuanced context might persist due to the model’s architecture.

### **Conclusion:**

The fine-tuned model exhibited improved accuracy on the test set, demonstrating the effectiveness of QLoRA in adapting to specific tasks. However, some failure cases persisted, indicating areas where the model still requires further improvement or more comprehensive training data.