LLM Assignment 3
Harshvardhan Singh, 2021052

Q1. Accuracy comparison between the pretrained and fine-tuned models on the test set:

Accuracy before fine-tuning: 54.0
Accuracy after fine-tuning: 87.0

Q2. Time taken to fine-tune the model using QLoRA.

Time taken to fine-tune the model (seconds): 2270.90

Q3. Total parameters in the model and the number of parameters fine-tuned.

Trainable Parameters: 20971520
Total Parameters: 1542364160
Percentage of Trainable Parameters: 1.36%

Q4. Resources used (e.g., hardware, memory) during fine-tuning.

Peak Memory Usage (GB): 3.24

Q5. Failure cases of the pre-trained model that were corrected by the fine-tuned model, as well as those that were not corrected. Provide possible explanations for both.

Number of cases that were corrected by the fine-tuned model: 34

Below are 5 such examples:
----------1----------
Premise: a woman in a black shirt looking at a bicycle.
Hypothesis: A woman dressed in black shops for a bicycle.
Label: 1
Pretrained Label: 0
Finetuned Label: 1
----------2----------
Premise: A group of people stand near and on a large black square on the ground with some yellow writing on it.
Hypothesis: a group of people wait
Label: 1
Pretrained Label: 0
Finetuned Label: 1
----------3----------
Premise: Two men in neon yellow shirts busily sawing a log in half.
Hypothesis: Two men are cutting wood to build a table.
Label: 1
Pretrained Label: 0
Finetuned Label: 1
----------4----------
Premise: A man is renovating a room.

Hypothesis: A man is using a hammer in a room.
Label: 1
Pretrained Label: 0
Finetuned Label: 1
----------5----------
Premise: An Ambulance is passing a man wearing a bandanna and girl.
Hypothesis: The man in the bandana is running after the ambulance
Label: 2
Pretrained Label: 1
Finetuned Label: 2

Number of cases that were not corrected by the fine-tuned model: 12

Below are 5 such examples:
----------1----------
Premise: This church choir sings to the masses as they sing joyous songs from the book at a church.
Hypothesis: The church has cracks in the ceiling.
Label: 1
Pretrained Label: 2
Finetuned Label: 2
----------2----------
Premise: Two men climbing on a wooden scaffold.
Hypothesis: Two sad men climbing on a wooden scaffold.
Label: 1
Pretrained Label: 2
Finetuned Label: 2
----------3----------
Premise: A woman is standing near three stores, two have beautiful artwork and the other store has Largo written on it.
Hypothesis: A woman standing on a street corner outside beside three different stores, two of which contain beautiful artwork and one with a Largo sign.
Label: 0
Pretrained Label: 1
Finetuned Label: 1
----------4----------
Premise: Military personnel are shopping
Hypothesis: Military personnel are in the mall.
Label: 1
Pretrained Label: 0
Finetuned Label: 0
----------5----------
Premise: An older gentleman wearing a hat is walking on crutches next to a busy street.
Hypothesis: A man with a walking stick is next to the street.
Label: 2
Pretrained Label: 1

Finetuned Label: 1

Explanation:

Language Nuance: The model can struggle with grasping subtle meanings and indirect implications in language.
Data Limitations: Limited variety in fine-tuning data may hinder the model's capacity to generalize well.
Architectural Restrictions: Performance may be constrained by the base architecture and the number of trainable parameters.
Pretraining Bias: Residual biases from pretraining can influence the effectiveness of fine-tuning.

References:
https://dassum.medium.com/fine-tune-large-language-model-llm-on-a-custom-dataset-with-qlora-fb60abdeba07,
https://www.kaggle.com/code/hari31416/downloading-file-and-directory-from-kaggle