LLM PA3 Report

Model Choice: Three models were chosen for inferencing with pre-training which were BART-LARGE-CNN, t5-large, Pegasus Xsum. Out of the three model choices BART-CNN-LARGE was chosen for further fine-tuning.

Dataset Preparation: The first step in the dataset preparation was to load the pretrained checkpoint from the BART model using the AutoTokenizer function of the transformers library. Once the tokenizer is loaded using this tokenizer, we tokenize both the dialogues and summaries with padding set to max_length. The last step is to divide the tokenized data into batches. Now the data is ready for processing.

Fine-Tuning Process: Once the data is ready the model is loaded using the AutoModelForSeq2SeqLM function of the transformers library. Rouge metric function is defined for evaluating the validation. The training arguments used for fine tuning this particular model is

```
training_args = Seq2SeqTrainingArguments(
    output_dir = '/content/drive/MyDrive/PA3',
    num_train_epochs=1,
    per_device_train_batch_size=2,
    per_device_eval_batch_size=2,
    warmup_steps=500,
    weight_decay=0.01,
    logging_dir='./logs',
    logging_steps=10,
    eval_strategy="epoch",
    learning_rate=2e-5,
    gradient_accumulation_steps=2,
    save_total_limit=2,
    predict_with_generate=True,
    fp16=True
)
```

Output_dir – The path for storing the finetuned model
num_train_epochs – The number of epochs for training the model. Since the BART Large is a very huge model and with the computational restrictions training was done only on 1 epoch
per_device_train_batch_size: Number of training examples per batch per device (GPU/CPU).
per_device_eval_batch_size: Number of evaluation examples per batch per device (GPU/CPU).
warmup_steps: Number of steps for the learning rate warmup to increase gradually from zero to the initial value.
weight_decay: The weight decay to apply to the optimizer, which helps in regularization and preventing overfitting.

logging_dir: Directory where the logs are saved.
logging_steps: Frequency (in steps) of logging training metrics and information.
eval_strategy: The evaluation strategy to use during training (e.g., “epoch” to evaluate at the end of each epoch).
learning_rate: The initial learning rate for the optimizer.
gradient_accumulation_steps: Number of steps to accumulate gradients before performing a backward/update pass.
save_total_limit: Maximum number of checkpoint files to keep; older ones are deleted.
predict_with_generate: Whether to use the model’s generate method to predict sequences during evaluation.
fp16: Whether to use 16-bit (half-precision) floating-point format for training, which can improve speed and reduce memory usage on compatible GPUs.

Summarization before Fine-Tuning

```
Original Conversation: Ralph: Have you ever heard of a company called
"Venque"?
Victoria: No. what is this?
Ralph: <file_other>
Ralph: Check it out
Charles: it's a Canadian company
Charles: They have very beautiful but expensive products.
Ralph: What do think about this bag?
Ralph: <file_other>
Charles: Very stylish
--------------------------------------------------------------------------
--------------------------------------------------------------------------
--
Original Summary: The Canadian company Venque has beautiful but expensive
products.
Summary from BART: Charles: They have very beautiful but expensive
products. Ralph: Have you ever heard of a company called "Venque"?
Victoria: No. what is this? Charles: It's a Canadian company called
Venque.
ROUGE Scores: {'rouge1': Score(precision=0.22580645161290322,
recall=0.7777777777777778, fmeasure=0.35000000000000003), 'rouge2':
Score(precision=0.13333333333333333, recall=0.5,
fmeasure=0.2105263157894737), 'rougeL':
Score(precision=0.12903225806451613, recall=0.444444444444444,
fmeasure=0.19999999999999998)}
**************************************************************************
**************************************************************************
**
Original Conversation: Elizabeth: that new restaurant at the corner of
chapel and college st finally opened
Aiden: oh really? what it it?
Elizabeth: it's a burger place
Aiden: ohh... i'm disappointed, i was hoping it would be something cooler
```

Elizabeth: such as?
Aiden: a vegan restaurant
Elizabeth: yeah, that would have been better
Aiden: having said that, would you like to go?
Elizabeth: sure, i'm not free tonight though
Aiden: tomorrow?
Elizabeth: that would be great
Aiden: and who knows? they might have great burgers!
Elizabeth: maybe even a vegan burger!!! lol
Aiden: hahaha that would be interesting
Elizabeth: can you pick me up?
Aiden: sure, what time?
Elizabeth: let's make it 9, i'll have a long day at work tomorrow
Aiden: wow, that's late
Elizabeth: i'm used to it
Aiden: ok, i'll see you tomorrow at 9, then!

--
Original Summary: The new burger restaurant at the corner of Chapel and College Street finally opened. Elizabeth and Aiden will go there tomorrow. Aiden will pick Elizabeth up at 9 o'clock.
Summary from BART: Elizabeth tells Aiden about a new burger place. Aiden is disappointed, thinking it would have been a vegan restaurant. Elizabeth asks Aiden to pick her up at 9pm tomorrow.
ROUGE Scores: {'rouge1': Score(precision=0.3793103448275862, recall=0.36666666666666664, fmeasure=0.3728813559322034), 'rouge2': Score(precision=0.07142857142857142, recall=0.06896551724137931, fmeasure=0.07017543859649124), 'rougeL': Score(precision=0.27586206896551724, recall=0.26666666666666666, fmeasure=0.2711864406779661)}
****
****
**
Original Conversation: Mike: R we meeting at 5 p.m. today?
Paul: Yes
Paul: Don't forget about it!
Paul: I know you
Mike: yes yes
Mike: I'll try :)

--
Original Summary: Mike reminds Paul that they are meeting at 5 pm today.
Summary from BART: Paul: Don't forget about it! Mike: R we meeting at 5 p.m. today? Paul: Yes. Mike: I'll try :)
ROUGE Scores: {'rouge1': Score(precision=0.2857142857142857, recall=0.5454545454545454, fmeasure=0.37499999999999994), 'rouge2': Score(precision=0.1, recall=0.2, fmeasure=0.13333333333333333), 'rougeL': Score(precision=0.23809523809523808, recall=0.45454545454545453, fmeasure=0.31249999999999994)}
****
****
**

Original Conversation: Jim: there are some cheap flights advertised
Jim: in January
Lane: What date
Jim: 15-24
Matt: I'm away on 15th & 16th
Matt: Can you find something later on?
Jim: there is one on 20th
Matt: That'd be fine
Matt: What time does it land?
Jim: 8:15pm
Matt: Ok, I can pick you up
Jim: Lane, are you okay with 20th
Lane: sure, no problem
Jim: thanks
Jim: so I will see all of you soon
Lane: can't wait
Jim: me too
Matt: bring me a few bottles of foreign goodies ;)
Jim: will do

--
Original Summary: Jim found cheap flights to visit Matt and Lane. He is flying in on the 20th, landing at 8:15pm. Matt will pick him up.
Summary from BART: Jim: There are some cheap flights advertised in January. Lane: I can pick you up at 8:15pm on 20th January. Jim: so I will see all of you soon Lane: can't wait
ROUGE Scores: {'rouge1': Score(precision=0.38235294117647056, recall=0.52, fmeasure=0.4406779661016949), 'rouge2': Score(precision=0.09090909090909091, recall=0.125, fmeasure=0.10526315789473685), 'rougeL': Score(precision=0.23529411764705882, recall=0.32, fmeasure=0.2711864406779661)}
****
****
**
Original Conversation: Paul: Hey Meg, how are your interviews going?
Meghan: Excellent! I've almost made it. There's only one interview left with the Hub Director 😊
Paul: Glad to hear that! It's cheered me right up!
Meghan: How's your work?
Paul: Not so bad, but I'm thinking of moving on and changing it for something much better
---

--
Original Summary: Meghan is having interviews and they are going well. Meghan is having just one more with the Hub Director. Paul is considering looking for a better job.
Summary from BART: Meghan: How's your work? Paul: Not so bad, but I'm thinking of moving on and changing it for something much better. Paul: Hey Meg, how are your interviews going?
ROUGE Scores: {'rouge1': Score(precision=0.25806451612903225, recall=0.2962962962962963, fmeasure=0.27586206896551724), 'rouge2':

```
Score(precision=0.0, recall=0.0, fmeasure=0.0), 'rougeL':
Score(precision=0.12903225806451613, recall=0.14814814814814814,
fmeasure=0.13793103448275862)}
```

Summarization after Fine-Tuning

```
Original Conversation: Ralph: Have you ever heard of a company called
"Venque"?
Victoria: No. what is this?
Ralph: <file_other>
Ralph: Check it out
Charles: it's a Canadian company
Charles: They have very beautiful but expensive products.
Ralph: What do think about this bag?
Ralph: <file_other>
Charles: Very stylish
--------------------------------------------------------------------------
--------------------------------------------------------------------------
--
```
Original Summary: The Canadian company Venque has beautiful but expensive products.
Summary from BART: Charles: They have very beautiful but expensive products. Ralph: Have you ever heard of a company called "Venque"? Victoria: No. what is this? Charles: It's a Canadian company called Venque.
```
ROUGE Scores: {'rouge1': Score(precision=0.22580645161290322,
recall=0.7777777777777778, fmeasure=0.35000000000000003), 'rouge2':
Score(precision=0.13333333333333333, recall=0.5,
fmeasure=0.2105263157894737), 'rougeL':
Score(precision=0.12903225806451613, recall=0.4444444444444444,
fmeasure=0.19999999999999998)}
**************************************************************************
**************************************************************************
**
Original Conversation: Elizabeth: that new restaurant at the corner of
chapel and college st finally opened
Aiden: oh really? what it it?
Elizabeth: it's a burger place
Aiden: ohh... i'm disappointed, i was hoping it would be something cooler
Elizabeth: such as?
Aiden: a vegan restaurant
Elizabeth: yeah, that would have been better
Aiden: having said that, would you like to go?
Elizabeth: sure, i'm not free tonight though
Aiden: tomorrow?
Elizabeth: that would be great
Aiden: and who knows? they might have great burgers!
Elizabeth: maybe even a vegan burger!!! lol
Aiden: hahaha that would be interesting
```

Elizabeth: can you pick me up?
Aiden: sure, what time?
Elizabeth: let's make it 9, i'll have a long day at work tomorrow
Aiden: wow, that's late
Elizabeth: i'm used to it
Aiden: ok, i'll see you tomorrow at 9, then!

--
Original Summary: The new burger restaurant at the corner of Chapel and College Street finally opened. Elizabeth and Aiden will go there tomorrow. Aiden will pick Elizabeth up at 9 o'clock.
Summary from BART: Elizabeth tells Aiden about a new burger place. Aiden is disappointed, thinking it would have been a vegan restaurant. Elizabeth asks Aiden to pick her up at 9pm tomorrow.
ROUGE Scores: {'rouge1': Score(precision=0.3793103448275862, recall=0.36666666666666664, fmeasure=0.3728813559322034), 'rouge2': Score(precision=0.07142857142857142, recall=0.06896551724137931, fmeasure=0.07017543859649124), 'rougeL': Score(precision=0.27586206896551724, recall=0.26666666666666666, fmeasure=0.2711864406779661)}
****
****
**
Original Conversation: Mike: R we meeting at 5 p.m. today?
Paul: Yes
Paul: Don't forget about it!
Paul: I know you
Mike: yes yes
Mike: I'll try :)

--
Original Summary: Mike reminds Paul that they are meeting at 5 pm today.
Summary from BART: Paul: Don't forget about it! Mike: R we meeting at 5 p.m. today? Paul: Yes. Mike: I'll try :)
ROUGE Scores: {'rouge1': Score(precision=0.2857142857142857, recall=0.5454545454545454, fmeasure=0.37499999999999994), 'rouge2': Score(precision=0.1, recall=0.2, fmeasure=0.13333333333333333), 'rougeL': Score(precision=0.23809523809523808, recall=0.45454545454545453, fmeasure=0.31249999999999994)}
****
****
**
Original Conversation: Jim: there are some cheap flights advertised
Jim: in January
Lane: What date
Jim: 15-24
Matt: I'm away on 15th & 16th
Matt: Can you find something later on?
Jim: there is one on 20th
Matt: That'd be fine
Matt: What time does it land?
Jim: 8:15pm

Matt: Ok, I can pick you up
Jim: Lane, are you okay with 20th
Lane: sure, no problem
Jim: thanks
Jim: so I will see all of you soon
Lane: can't wait
Jim: me too
Matt: bring me a few bottles of foreign goodies ;)
Jim: will do

--
Original Summary: Jim found cheap flights to visit Matt and Lane. He is flying in on the 20th, landing at 8:15pm. Matt will pick him up.
Summary from BART: Jim: There are some cheap flights advertised in January. Lane: I can pick you up at 8:15pm on 20th January. Jim: so I will see all of you soon Lane: can't wait
ROUGE Scores: {'rouge1': Score(precision=0.38235294117647056, recall=0.52, fmeasure=0.4406779661016949), 'rouge2': Score(precision=0.09090909090909091, recall=0.125, fmeasure=0.10526315789473685), 'rougeL': Score(precision=0.23529411764705882, recall=0.32, fmeasure=0.2711864406779661)}
****
****
**
Original Conversation: Paul: Hey Meg, how are your interviews going?
Meghan: Excellent! I've almost made it. There's only one interview left with the Hub Director 😊
Paul: Glad to hear that! It's cheered me right up!
Meghan: How's your work?
Paul: Not so bad, but I'm thinking of moving on and changing it for something much better
---

--
Original Summary: Meghan is having interviews and they are going well. Meghan is having just one more with the Hub Director. Paul is considering looking for a better job.
Summary from BART: Meghan: How's your work? Paul: Not so bad, but I'm thinking of moving on and changing it for something much better. Paul: Hey Meg, how are your interviews going?
ROUGE Scores: {'rouge1': Score(precision=0.25806451612903225, recall=0.2962962962962963, fmeasure=0.27586206896551724), 'rouge2': Score(precision=0.0, recall=0.0, fmeasure=0.0), 'rougeL': Score(precision=0.12903225806451613, recall=0.14814814814814814, fmeasure=0.13793103448275862)}

Rouge Scores before Fine-Tuning

{'rouge1': AggregateScore(low=Score(precision=0.2316205533596838, recall=0.36918518518518517, fmeasure=0.30422372227579564), mid=Score(precision=0.28501848782353695, recall=0.5012390572390573, fmeasure=0.3477467052394728), high=Score(precision=0.33740323997336624, recall=0.6797575757575757, fmeasure=0.38693989071038243)),

'rouge2': AggregateScore(low=Score(precision=0.03428571428571429, recall=0.05, fmeasure=0.04067796610169492), mid=Score(precision=0.07469696969696969, recall=0.17629310344827587, fmeasure=0.09889830508474577), high=Score(precision=0.10547619047619047, recall=0.33999999999999997, fmeasure=0.15389830508474578)),

'rougeL': AggregateScore(low=Score(precision=0.14044795783926217, recall=0.22992592592592592, fmeasure=0.17055425448868072), mid=Score(precision=0.18802045702466458, recall=0.32676094276094275, fmeasure=0.22850346696055474), high=Score(precision=0.23465092722455974, recall=0.4235959595959596, fmeasure=0.2750241080038573)),

'rougeLsum': AggregateScore(low=Score(precision=0.14044795783926217, recall=0.2296592592592595, fmeasure=0.1767596546815449), mid=Score(precision=0.18802045702466458, recall=0.32676094276094275, fmeasure=0.22850346696055474), high=Score(precision=0.23462677263518777, recall=0.4235959595959596, fmeasure=0.2750241080038573))}

Rouge Scores after Fine-Tuning

{'rouge1': AggregateScore(low=Score(precision=0.1, recall=0.05703703703703704, fmeasure=0.0731457800511509), mid=Score(precision=0.2638888888888889, recall=0.16176430976430978, fmeasure=0.1958178448284788), high=Score(precision=0.4288888888888889, recall=0.28218181818181814, fmeasure=0.3295546558704453)),

'rouge2': AggregateScore(low=Score(precision=0.0, recall=0.0, fmeasure=0.0), mid=Score(precision=0.06547619047619048, recall=0.047692307692307694, fmeasure=0.05505882352941176), high=Score(precision=0.17976190476190473, recall=0.1276923076923077, fmeasure=0.1491764705882353)),

'rougeL': AggregateScore(low=Score(precision=0.06944444444444445, recall=0.0399555555555556, fmeasure=0.04660633484162896), mid=Score(precision=0.1934444444444445, recall=0.12428282828282829, fmeasure=0.14820816550524454), high=Score(precision=0.35, recall=0.2537373737373737, fmeasure=0.29355229505990044)),

'rougeLsum': AggregateScore(low=Score(precision=0.07193611111111116, recall=0.0399555555555556, fmeasure=0.047047511312217204), mid=Score(precision=0.1934444444444445, recall=0.12428282828282826, fmeasure=0.14820816550524452), high=Score(precision=0.34900000000000003, recall=0.24751515151515152, fmeasure=0.29309980637211763))}

Analysis

- Before Fine tuning the model was just outputting the main dialogues and there was no coherence in the summaries
- Just with one epoch the model's summarization did improve a little bit but still outputs the dialogues as the summarization except that this time it catches the correct conversations to match the original summaries.
- When comparing the rouge scores from before fine tuning to after fine tuning we can clearly see an increase in the highs of all the rouge metrics