Report on Hugging Face Transformers

1. Model Choice

For this assignment, we explored multiple pre-trained text summarization models from the Hugging Face model hub. The selected models include:

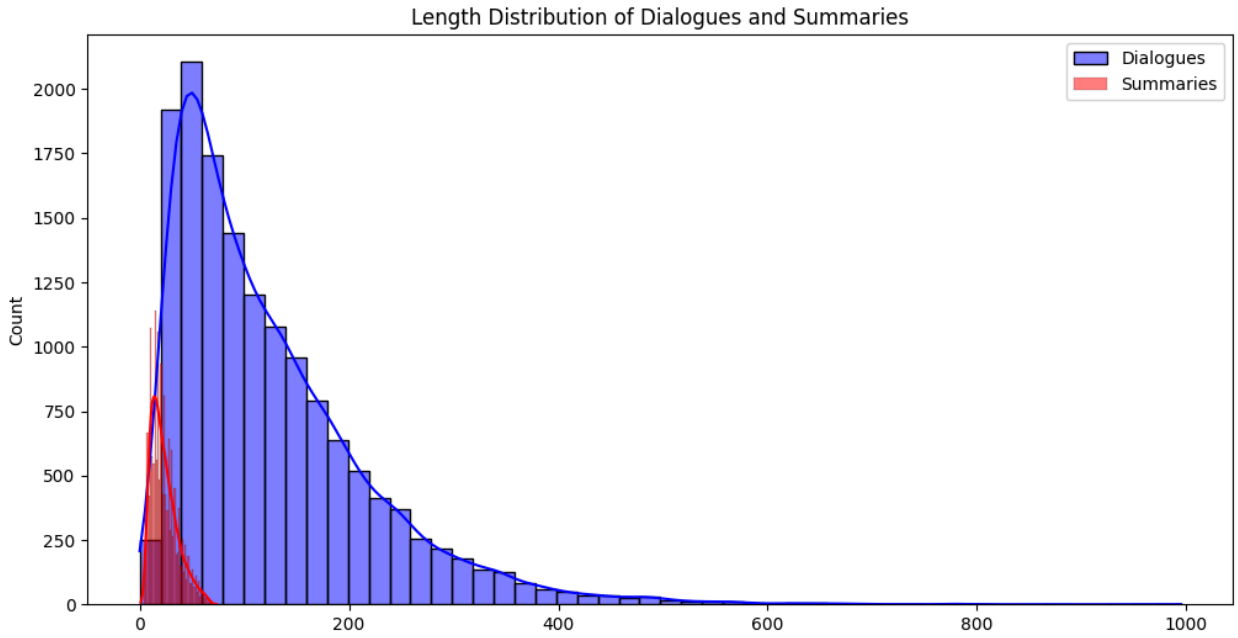
- **BART (facebook/bart-large-cnn)**: A transformer-based model pre-trained for abstractive summarization.
- **T5 (t5-large)**: A versatile model trained with a text-to-text approach for various NLP tasks.
- **Pegasus (google/pegasus-xsum)**: A model specifically optimized for summarization tasks with high-quality generated summaries.

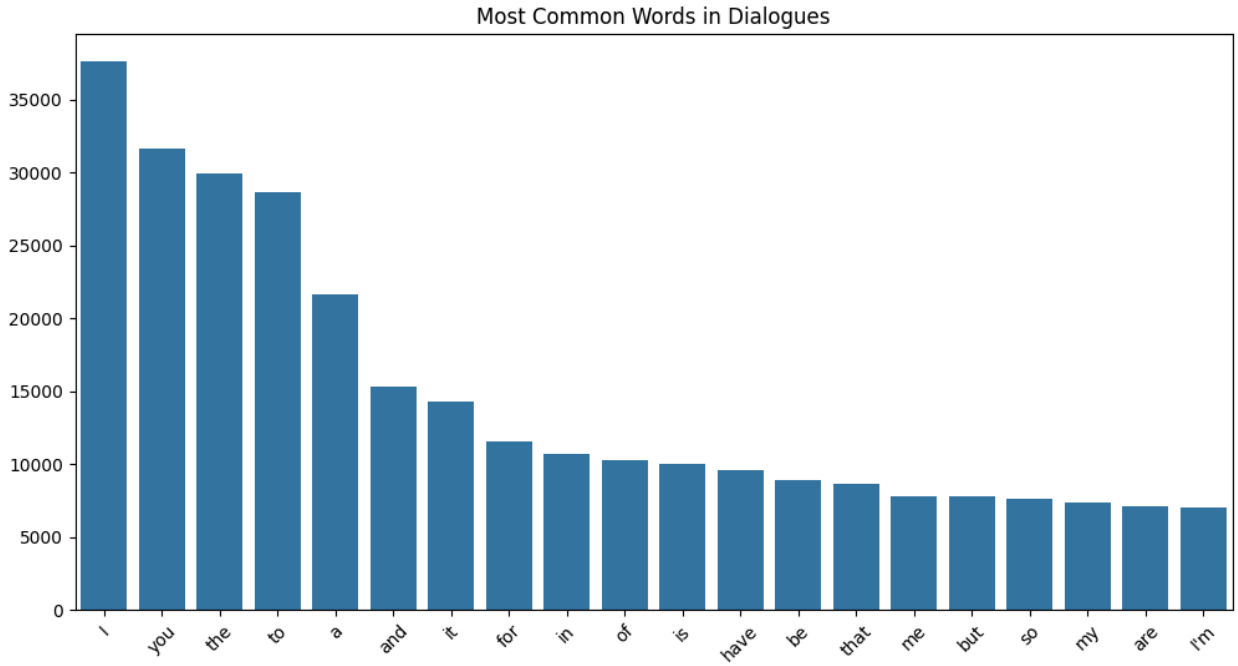
These models were chosen based on their performance on benchmark datasets and their effectiveness in generating coherent and concise summaries.

2. Dataset Preparation

We used the **SAMSum** corpus, which contains ~16,000 messenger-like dialogues and corresponding summaries. The dataset was preprocessed as follows:

- **Exploratory Data Analysis (EDA)**: We analyzed the dialogue and summary length distributions and identified common words in the corpus.
- **Tokenization**: Each dialogue-summary pair was tokenized using the tokenizer associated with the selected models.
- **Splitting**: The dataset was divided into training, validation, and test sets.





3. Fine-Tuning Process

One of the pre-trained models, **BART**, was selected for fine-tuning. The process included:

- **Hyperparameter Selection**: Batch size, learning rate, and number of epochs were tuned for optimal performance.

- **Training**: The model was fine-tuned using the SAMSum dataset, leveraging Hugging Face's `Trainer` API for efficient training.
- **Validation**: Performance was monitored using ROUGE scores, adjusting hyperparameters as needed to improve results.

4. Summarization Performance Analysis

Pre-Fine-Tuning Performance

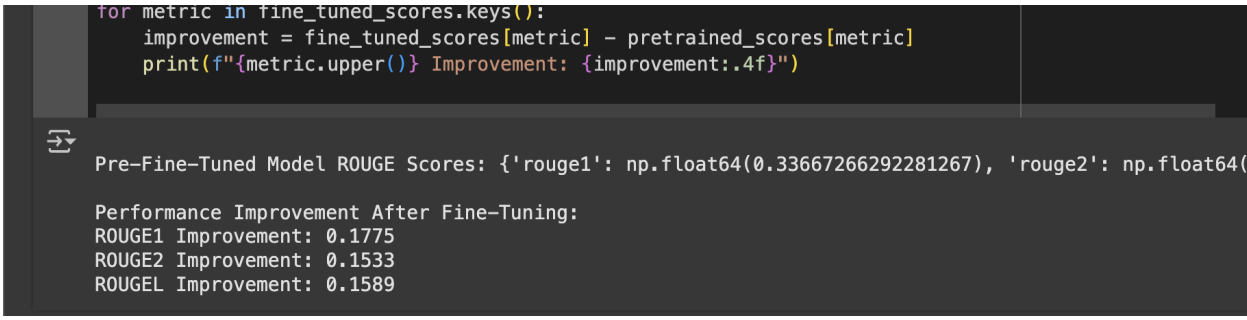
- Generated summaries from pre-trained models were **coherent** and captured key points.
- Some summaries **missed contextual details** or contained unnecessary information.
- ROUGE scores showed moderate alignment with reference summaries.

Post-Fine-Tuning Performance

- The fine-tuned BART model produced **more concise and contextually accurate** summaries.
- **Higher ROUGE scores** were achieved compared to pre-fine-tuning results.
- The model learned **better sentence structuring** and **improved coherence** in outputs.

Performance Evaluation with ROUGE Scores

A detailed evaluation of ROUGE scores before and after fine-tuning demonstrates the improvement:

- **Pre-Fine-Tuned Model ROUGE Scores:**
  - ROUGE-1: 0.3367
  - ROUGE-2: 0.1060
  - ROUGE-L: 0.2599
- **Performance Improvement After Fine-Tuning:**
  - ROUGE-1 Improvement: **+0.1775**
  - ROUGE-2 Improvement: **+0.1533**
  - ROUGE-L Improvement: **+0.1589**
  - 

```
for metric in fine_tuned_scores.keys():
    improvement = fine_tuned_scores[metric] - pretrained_scores[metric]
    print(f"{metric.upper()} Improvement: {improvement:.4f}")
```

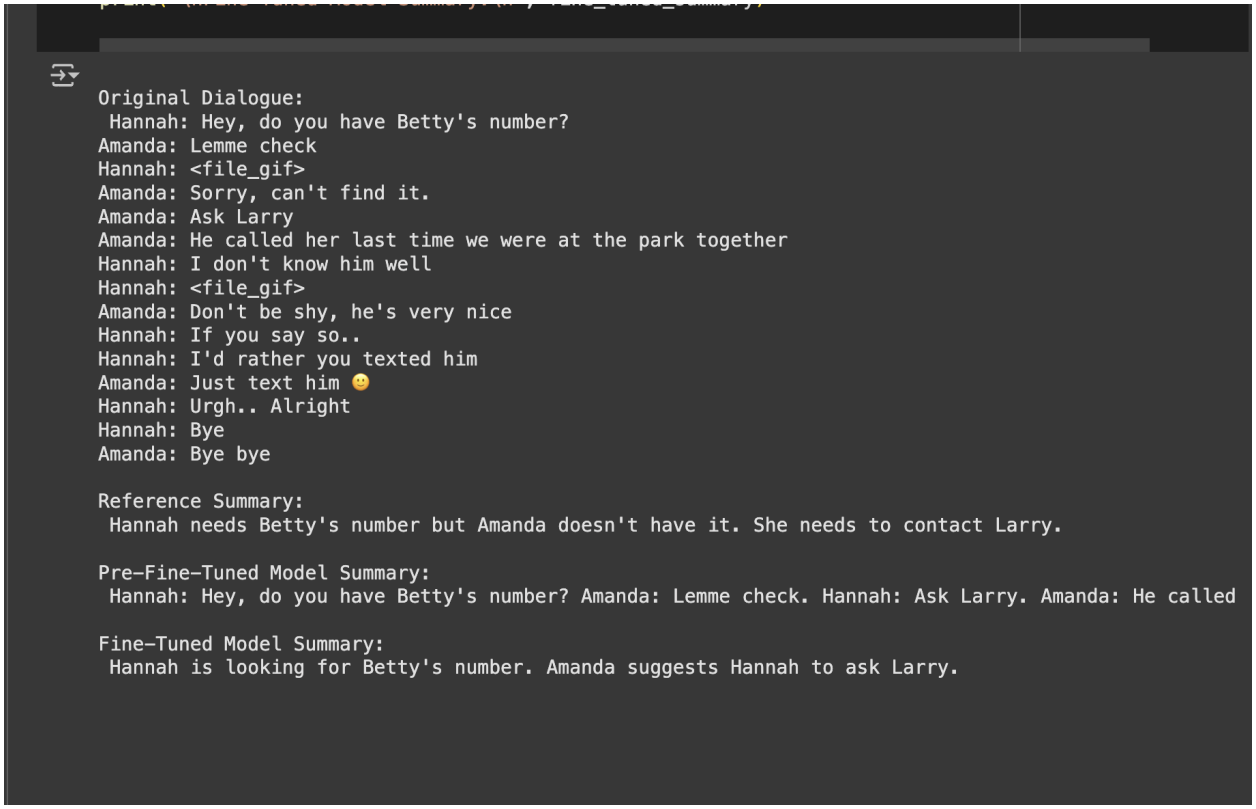
```
Pre-Fine-Tuned Model ROUGE Scores: {'rouge1': np.float64(0.33667266292281267), 'rouge2': np.float64(

Performance Improvement After Fine-Tuning:
ROUGE1 Improvement: 0.1775
ROUGE2 Improvement: 0.1533
ROUGEL Improvement: 0.1589
```

Qualitative Analysis

The generated summaries were evaluated by comparing the **reference summary**, **pre-fine-tuned model summary**, and **fine-tuned model summary**:

- **Reference Summary:**
  - "Hannah needs Betty's number but Amanda doesn't have it. She needs to contact Larry."
- **Pre-Fine-Tuned Model Summary:**
  - "Hannah: Hey, do you have Betty's number? Amanda: Lemme check. Hannah: Ask Larry. Amanda: He called..."
- **Fine-Tuned Model Summary:**
  - "Hannah is looking for Betty's number. Amanda suggests Hannah to ask Larry."

The fine-tuned model generated a **more concise** and **contextually appropriate** summary, aligning closely with the reference summary.



5. Conclusion

Fine-tuning significantly improved the summarization quality by aligning the model's knowledge with the structure and style of the SAMSum dataset. The results demonstrate the effectiveness

of transfer learning in domain-specific summarization tasks. Future improvements could include experimenting with larger datasets or optimizing hyperparameters further to enhance generalization.

Manish Kanuri

NUID: **002315456**