Assignment 4: Transformer

Andrew Nguyen (AXN210071)

CS 4372.501

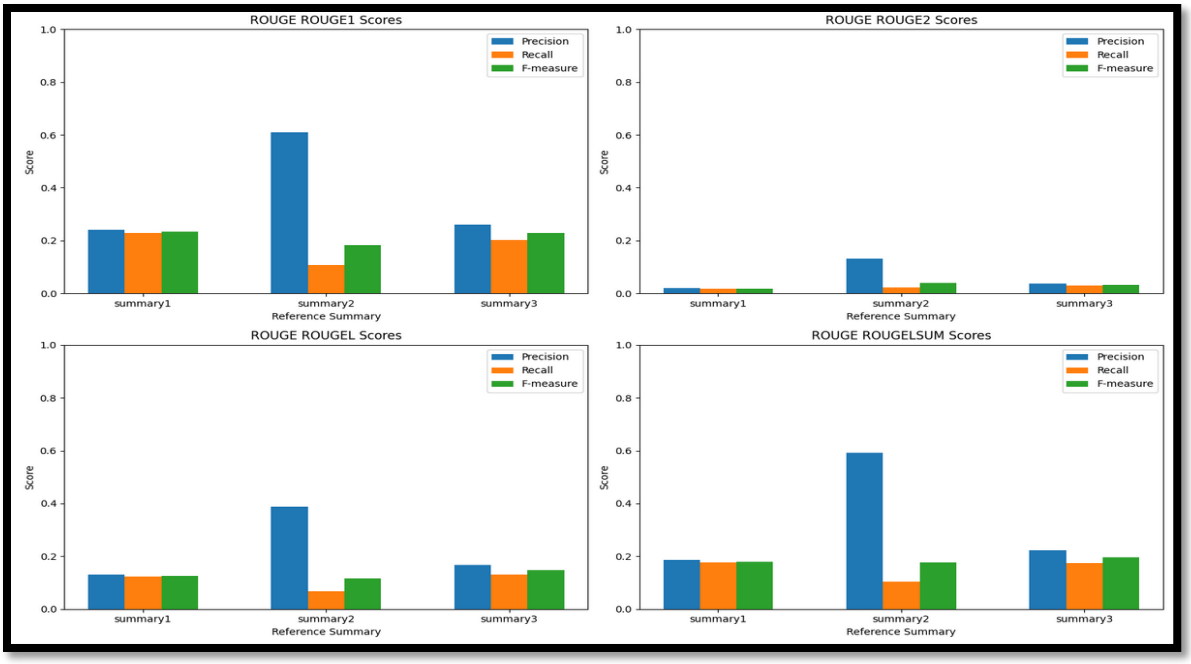
Link to Colab:
https://colab.research.google.com/drive/1Bn08EJPrgkWNqQhM03Is00RLGrdvZAuC#scrollTo=jwQX90GmjD0u

Link to GitHub: https://github.com/AndrewN20/CS4372.501Assignment4Transformer

Link to model: https://huggingface.co/google-t5/t5-small

I decided for this assignment to do text summarization of the Monkey's Paw (https://www.gutenberg.org/ebooks/12122). I decided to use the "google-t5/t5-small" model for my text summarization assignment. I first downloaded the short story as a text file and removed the copyright/legal information in the text. Then I loaded the model in a pipeline and calculated the ROUGE Scores, according to this article: https://thepythoncode.com/article/calculate-rouge-score-in-python.

Each ROUGE score is calculated by comparing the generated summary from the model and then comparing three different summaries: a short AI summary (Summary 1), the summary from Wikipedia (Summary 2), and my own summary (Summary 3). Looking at the results of the rogue score, we see that overall, the generated summary scores are low in the precision, recall, and F-measure for all types of ROGUE scores. Though note that for Summary 2, the precision is high, meaning that the generated summary is mostly accurate compared to the Wikipedia summary. I believe the reason why precision is unusually high for Summary 2 is because in the model's research paper, it mentions that Wikipedia was used for pre-training. Overall, we can say that our generated summary text doesn't perform well.



I investigated the documentation to see if I can do any fine-tuning. But I couldn't find anything related to fine-tuning, at least on my level. When I checked on how their model could be fine-tuned, it required the use of running a TPU (which I believe requires a higher tier of Google Colab) and set up external information, like a Google Cloud Bucket (https://colab.research.google.com/github/google-research/text-to-text-transfer-transformer/blob/main/notebooks/t5-trivia.ipynb#scrollTo=i_-7qYemnEHl). Without the ability to properly fine-tune the model properly without any additional setup, our ROGUE score metrics are relatively low. Overall, the model needs improvement in terms of trying to improve ROUGE score if we had access to fine-tuning.