

Quoth the Bot

Goal of the Project:

My goal was to train a transformer-based model to convert modern English into a Shakespearian style.

Resources used:

I used Pytorch to run and train the model.

For this project I started by using the Hugging Face transformer library, specifically the T5 text-to-text transfer transformer.

For training data, I used the paired modern vs original Shakespeare dataset from this similar project on Github: https://github.com/Jmkernes/Shakespeare-Translator/tree/main/data_clean

Physically speaking I was using a Quadro P1000 4gb graphics card to train the model.

Process and Experimentation:

After a lot of documentation reading, I was able to get the model training and saving with checkpoints (in case the computer crashed). This process was extremely slow however, so I tried a bunch of different things to try and speed it up.

One thing I attempted to do was to use another P1000 card in the computer and have both graphics cards used for training at the same time. I sadly could not get them to work in tandem, as I would constantly get out of memory errors on the cards even when reducing the intensity of the training session by an extreme amount. Unfortunately, as I did not have any other graphics cards available, I am not sure if that was a symptom of my model, the computer, or the graphics cards themselves.

The main thing that helped speed up the training was playing with the batch size. I originally had the batch size set to 16 as that was recommended to me. This would result in training for 100 epochs over the whole dataset taking around 36 hours for my system. I found this extremely inconvenient as I would not have a consistent time at which I could restart the process each day. After looking for advice and the specifics of my graphics card I found tried reducing the batch size to 8 and I found it increased the speed by a significant amount, dropping to around 22 hours.

After I cut down the training time I trained the model for roughly 200 hours to get it to its current checkpoint.

Conclusion:

In the end the model has taken quite well to the Shakespeare style with some few issues. The

main issue being that sometimes it simply won't change anything from the input. I am not sure if that is because the model needs more training or if it simply doesn't think the sentence needs to be changed. The other issue comes from the data set, namely the lack of capitalization. This is a minor problem, but since the training and testing data have no capital letters, the responses from my model also have none.