**Phase One – Baseline Testing**

This phase of the project included having each member of the team run different models from HuggingFace in order to establish a baseline model which could later be fine-tuned.

**Longformer model on the Quail Dataset**

The longformer was chosen because it works well with longer documents.

This baseline test has been completed. The performance of this model was:

Loss = 0.8903

Accuracy = 0.6682

Epoch = 1.99

The model can be found here:

<https://colab.research.google.com/drive/1TguhpUu2Eqfos6_IIGWhfn94FK-5tpVP?authuser=1>

**Roberta on the Quail Dataset**

For this baseline test, the Quail Dataset was converted into the Swag dataset format then ran on the Swag Model. This was done because we feel our data better aligns with the data of the Swag model, so we wanted to see how well the baseline performed on the new format.

This model is still running on Google Collab and has been running for over 100 hours. The code can be found here:

<https://colab.research.google.com/drive/1lLzcEaj1spbiXCJKprF-a7e4fSFTfkD_?authuser=1>

**Robert on the Swag Dataset**

The Swag baseline was ran on the Swag dataset due to one of the synthetically created datasets having a shorter context than the others. The team wanted to have an idea on how well this dataset would perform on its own so we could understand how it affects the overall performance.

This baseline test has been completed. The performance of this model was:

Loss = 1.018

Accuracy = 0.7937

Epochs = 3

The model can be found here:

<https://colab.research.google.com/github/ViktorAlm/notebooks/blob/master/MPC_GPU_Demo_for_TF_and_PT.ipynb?authuser=1#scrollTo=Q75NAj12wT7T>