Over the past week we have spent a large amount of time working on data aggregation for our experiment. We looked into both what types of data and data sources we wanted as well as ways in collecting mass quantities of data to best test our experiment. We decided that we would include about 100 documents from each category of bad, medium, and great; for great, we would use the introduction and conclusion from research level papers. For medium, we would use high school to college application level papers/essays. And for bad, we would use elementary level writing. In order to collect the mass quantities of documents, we would find an initial source and then create a python script that scalps the online source and saves all of the documents into individual word documents. We also spent a bit of time refining our experimental design from the initial small sample test to the full scale. We decided we would run multiple independent runs of the experiment by sample from our database (without replacement) until we have covered every document. Finally we refined our prompts and parameters for the LLM. This sets us up to have our data ready to run our experiments with enough variety in terms of the essay level as well as having a solid overall sample size, which will allow us to better generalize our results to this situation.