My data is a combination of real human answers for an in-class quiz and AI Generated answers created using ChatGPT. The AI was given the same question series that humans were, and answers were re-generated to match the number of human answers for a given question. This data was then organized in a .tsv file with data on the sentence number, quiz question, and whether or not the answer was produced by a human or by AI. Because ChatGPT is a relatively new and ever-changing software, there is currently no extensive body of research comparing human and AI-generated answers. Most recent research available seems to focus on ChatGPT’s merits in writing or on its theoretical societal implications, but there are few studies on ChatGPT’s efficacy in a classroom setting. Through my project, I aim to explore the ability of ChatGPT to mimic real human answers when utilized by students.

In an effort to build my corpus, I utilized automated POS tagging. To do this, I employed a Python script utilizing spaCy and fed the script each individual sentence, resulting in a semi-automated tagging system. I then personally analyzed these tags to ensure there were minimal errors within the dataset. I chose to use the column format for my POS tagging to fit with the format given by spaCy. This created a slight challenge in organizing the data, since I needed to fit information from the original .tsv file into a simple .txt format. To solve this problem, I elected to label each sentence with its relevant number in the .tsv file and utilized parentheses to identify the question number and answer source. I believe that this was the best way to ensure the data was the most readable without requiring the user to cross-reference POS tags with the .tsv file. I also attempted to utilize adequate whitespace in an effort to make the data more readable and to avoid the confusion of adjacent sentences. In these ways, I believe I was able to make my POS tagging as readable as possible.

During a cursory glance at my POS tagging, I noticed that there seemed to be a specific pattern followed by nearly all AI-generated answers. Many ChatGPT answers seemed to be fairly stiff in structure, with exceedingly similar grammar. This was most prominent in specific questions, such as question 5, where every single AI generated answer began with “corpora of spontaneous spoken language are relatively rare.” Even when questions elicited relatively different responses by AI, sentence structure was still stiff enough that POS tags were similar across multiple answers. As a result, I began to question whether or not AI influence could be identified specifically by the syntactic structure of the answer. I believe it would be interesting to investigate if there is a preference for certain phrase structures or parts of speech in AI vs. human answers, which could be done by comparing the frequency of various parts of speech in the data. It would also be interesting to examine the syntactic structure of different answers to see the difference in patterns between human and AI.