CONNECTICUT COLLEGE   **COM110: HW3**

For this homework you will write a response and upload your pdf result. Responses can be handwritten or typed but you will need to make sure to run a pdf conversion on either. For hand written you will get the best/easiest result by using black ink on white paper and placing the sheet on a flat surface before using your phone to take a picture. From there you should be able to convert to pdf with an app if it is not built in already. For typed responses using Microsoft Word (all students get this for free via IS) you can simply do a *save as* and then pick pdf. Moodle allows TAs and your instructor to view the pdf rendering without downloading which is an efficient way to grade and provide feedback. Each response should be one well thought paragraph (3-5 sentences).

[1 20pts] Chat bots have been very popular this year, but often users are not aware of how the responses are generated. Looking at ChatGPT 3 as an example, we will attempt to use what we have learned in COM110 so far to better understand how this system functions. Navigate to: https://chat.openai.com/ and sign up using your conncoll.edu gmail. Then you will have access to the 3.5 API window which askes you to: **Send a Message**. Play around with sending 3-5 different messages for part [1]. You can simply write your inputs here for this problem.

[2 20 pts] Think carefully about what type of data you would be sending if this was a python program. What are you sending in terms of data type?

[3 20 pts] Think carefully about what type of data you are getting as a response. What are you receiving in terms of data type? Hit **Regenerate** and compare the response.

[4 20 pts] Can you give prompts to the input that will yield a pure numerical result? If so, please detail your message and what you received back here.

[5 20 pts] The last openai paper where they shared the details of the model were in Chat-GPT3 which can be located here: https://arxiv.org/pdf/2005.14165.pdf. Download this paper and look at Figure 2.2 on page 9 which shows the amount of text used to train it. Now think about the total amount of storage needed to put the training set. How big of a HDD or SDD do you need, if every token was 8-bytes? Hint multiply the total tokens for all data in the figure 2.2 by 8 for the total number of bytes and then try to reduce using the fact that 1024 bytes is a KB and 1024 KB is a MB and 1024 MB is a GB.