1. What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.

   Name: Toby Liang
   NetId: tobyzl2
   Captain: Toby Liang

2. What is your free topic? Please give a detailed description. What is the task? Why is it important or interesting? What is your planned approach? What tools, systems or datasets are involved? What is the expected outcome? How are you going to evaluate your work?

   For my free topic, I will be creating a recommendation system for research papers. In this task, the user will submit a research paper that interests them and the recommendation system will find a set of relevant papers from a corpus of research papers that is most relevant. This task is important because users who perform literature reviews need to find papers that are similar in topic and researchers can sometimes realize that a paper is not so relevant in the middle of reading a paper. Thus, this tool can save lots of time for researchers as it can automatically compare the content between different papers.

   My planned approach is to create a dataset using the arxiv api where the data will be the abstracts of the papers and the labels will be the arxiv category. Using this dataset, I will then train a classifier that takes in the paper abstract and predicts an arxiv category. At inference, papers with the same category will be considered for recommendation and ranked using term overlap. To write my classifier, I will utilize the Pytorch library. The expected outcome is a ranked listing of relevant papers that are predicted to be in the same arxiv category. To evaluate my work, I will create a test set to evaluate my classification model using precision, recall, and F1 score. Finally, I will create an API for this tool using Flask so that developers can have easy access to this tool.

3. Which programming language do you plan to use?

   I plan to use Python since many of the APIs and libraries I plan to use are written for Python.

4. Please justify that the workload of your topic is at least 20*N hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.

   Manual Data Collection: 4 Hours
   Writing Model and Training: 6 Hours
   Writing Keyword Matching and Ranking: 4 Hours
   System Evaluations: 3 Hours

Writing Flask API: 3 Hours