Group members and contact info:

- Jake Cumberland, NetId: jakepc3, email: jakepc3@illinois.edu
- Miles Cheng, Netid: milesc4, email: milesc4@illinois.edu
- Jaehong Lim, NetId: jaehong6, email: jaehong6@illinois.edu

Group Leader: Miles Cheng

1. 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?

The free topic we want to investigate is sentiment analysis, specifically on instagram posts based on a posts comment section. We want to examine the comments and categorize the post as well received, neutral, or bad based on the comments a post receives. Social media is used not only for social purposes, but also for influencers who seek their livelihood on social media platforms; a sentiment analysis tool could be helpful in guiding influencers towards making content their users enjoy more, thus boosting revenue. We plan on investigating the python NLTK library for sentiment analysis and Scikit Learn, Pandas, and Numpy to make a model which can predict how well a post was received based on the sentiment analysis. We will also be constructing an instagram post dataset to train our AI model on. The expected outcome is that we have a tool that can accurately predict whether an instagram post was well received. We will evaluate the tool based on how accurately it can predict reception of instagram posts from our created dataset.

2. Which programming language do you plan to use?
We plan to use python.

3. 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.

The project we are proposing has three main pillars: Creating a sentiment analysis tool meant to predict a posts reception based on comments, building a tool to scrape instagram posts online and process them into a format we can pass through our sentiment analysis tool, and creating a dataset to test and train our sentiment analysis tool on. Simply pursuing these goals will likely net us the 60 hour justification, but there will be many steps which will further increase our workload. This includes debugging the project, learning the libraries required to build these tools, as well as modifying functions to fit our specific needs. If there is extra time, some additions that we can add to the project could possibly include visualizations on the comment to like ratio, sentiment analysis results, and overall posts sentiment analysis results of a profile.