Final Project: Proposal

#### Song Lin (sla475, 301550328), Shlok Nangia (sna101, 301550257), Jialiang Guo (jga169, 301558771)

### **1. List 3 questions that you intend to answer (1 point)**

* Can we develop a classification model that can accurately distinguish between fake and real news on Twitter?
* Is it possible to identify trends in fake news on Twitter, including which type of accounts are more likely to produce these fake news?
* Can we determine the source of these fake news articles and how they spread?
* Can we group the fake articles into categories (such as political, sporting, social, etc.) to determine the most frequent ones?

### **2. List all the datasets you intend to use (1 point)**

Following are a few datasets we plan to use:

* [FakeNewsNet](https://github.com/KaiDMML/FakeNewsNet/tree/master/dataset)
* [liar-politifact dataset](https://paperswithcode.com/dataset/liar)
* [Fake News Classification - NLP](https://www.kaggle.com/code/lykin22/fake-news-classification-nlp/data)

### **3. Give us a rough idea on how you plan to use the datasets to answer these questions. (2 points)**

* Data Collection: Direct download of the above mentioned datasets
* Data Exploration: Conduct EDA on the datasets to understand the structure and characteristics of the data.
* Data Cleaning: Remove irrelevant columns, filter out low-quality records, and normalize the data if needed.
* Data Integration: As we are using multiple data sources we may have to merge all the datasets.
* Data Analysis: Use NLP, graph analysis, clustering, and classification techniques to analyze the data.
* Data Product: Develop a classification model that can accurately distinguish between fake and real news on Twitter. And create a report on the findings, including trends in fake news on Twitter and the genres of accounts that are more likely to produce it.

### **4. Think about that once your project is complete, what impacts it can make. Pick up the greatest one and write it down. (1 point)**

Recent years have witnessed fake news becoming a major problem, particularly on social media platforms like Twitter and Facebook. Our project aims to **identify and classify these fake news and articles**, **to improve media literacy**, protect public from the effects of fake news, and support easier fact-checking.