# Twitter Censorship Prediction Model

Week 5 Assignment - Group Project Proposal

**1. Motivation:**

The topic of censorship and free speech on the internet has flooded the news in recent years and rules regarding what should and shouldn't be censored is often hidden behind company policy and riddled with ambiguity.

Our objective is to create a machine learning model that can predict whether a tweet will likely be censored by twitter or not.

This would be a useful tool to determine what tweets twitter determines as acceptable, and to enable users to gauge the likelihood that a particular tweet will be censored.

**2. Dataset:**

We intend to collect our data from both the Twitter API, and a dataset called Twitter Stream Grab, which stores 1% of all tweets for each month up until August of 2021. For more recent tweets we will need to scrape it from the API ourselves.

The Twitter dataset has Tweet objects which include information such as the user, the contents and more importantly if the tweet has been censored and in what countries it has been censored.

**3. Method:**

This is a classification problem and we are planning to apply existing learning algorithms to the problem. We will need to tokenize the text so that it can be processed by the model.
We are planning to use the Transformers library from Huggingface.co to interface with various existing models, such as BART or ALBERT. These pretrained models will enable us to quickly tokenize the text from the tweets and use features such as device, location, user etc to ultimately train the model.

**4. Intended experiments/testing:**

We plan to use a given month of data from the Twitter Stream Grab as our testing data, and use confusion matrices to demonstrate the efficacy of the model.