```mermaid
flowchart TB
```

Write a script that can scapes all the comments on the top threads of r/WallStreetBet that mention "GameStop" or "GME".

Write a script that takes in a stock name as a parameter, and replaces all mentions of that stock name with a generic term.

Design a HIT that asks a Turker to determine if a phrase scrapped from the subreddit feels positively or negatively about GameStop

Write a script that splits all the scraped comments into manageable phrases (ex. At most a couple of sentences)

Qualification 1: Workers must be from America since they make up the majority of the users.

Qualification 2: Workers will be from younger age ranges.

Have 5 Turkers evaluate each comment scraped in parallel, as in having 5 completions per task for our HIT. This ensures the multiple opinions will be given on each phrase.

Turkers will be given a small text box that acts as reference for common reddit slang and its meanings

Workers will be presented a short survey in the beginning that asks about their knowledge/experiences with stock trading, the amount of time they spend on the reddit, and their attitudes towards stock trading/reddit/ This will give us a better gauge of the motivations of users and the demographics

Dispatch the HIT onto MTurk and take the resulting CSV into consideration for the next step of the project

Quality Module

Parse through the results we get label each of those phrases by the plurality of the given responses.

Run Weighted Majority Vote on the parsed data from Mechanical Turk, with weights be assigned based on the quality control responses of workers

Throw out all results that were submitted in less than the average amount of time it takes to reach a comment.

Aggregation Module

Iterate through the newly labeled data and determine the overall sentiment of r/WallStreetBets towards a particular stock by finding the most cited sentiment throughout all scrapped comments.

Format all of our data so that we are left with a CSV containing a set of phrases that are each labelled with a connotation that is either "Positive", "Negative", or "Neutral"

Feed this labeled data into a Categorical Classifier fast.ai neural network. Once the network is trained, it should be able to make a guess if an inputted string feels positively about

Using our scraping script, scrape all posts pertaining to the requested stock. Then using our mention replacer script, replace all mentions of those stocks with our predetermined generic stock.

Using our cleaned dataset of Reddit posts, feed all of them into the neural network. This should output the percentages of Reddit's feelings about this stock.

Final output should read something along the lines of "about __% of r/WallStreetBets feels positively about this stock".