CS492 Winter 2024
Final Project Milestone #1 - Project Proposal
February 16, 2024
Alex Hejmej [20779600], Tailai Wang [20816387], Isaiah Witzke [20845973]

Interactive Storytelling Website about AI Bot Toxicity on Social Media Platforms

Our group is curious about how many comment sections, forums, and online communities are often plagued by content generated by non-humans. Often referred to as Social Bots, AI Bots, or just 'Bots', many social media platforms are known for having problems with spam, phishing, and harmful content being shared by these accounts. When Web3 and NFTs were extremely popular, we noticed that Twitter was overrun with obviously fake accounts encouraging users to buy into their pyramid scheme. Around the same time, we noticed how some bots managed to draw extra engagement to extreme posts on Facebook and Reddit concerning vaccine propaganda and conspiracy theories. Our wish is to explore how successful these bots are at creating artificial engagement on social media, and whether or not platforms are able to moderate them at acceptable rates. We wish to dissect a few questions of interest:

1. Where do bots come from / who is making these bots?
2. What are bots' common message(s)/intention(s)? Are bots effective at drawing attention to their message(s)?
3. Are bots able to manipulate a social media user into engaging with their content?
4. What are social media platforms doing to moderate & shut down bots? Are they successful?

Methodology: 3 Phase Approach

1. **Before March**: Analysis and synthesis of existing literature / content
   a. A cursory search shows a fair amount of academic papers who discuss the prevalence of bots on mainstream platforms (like this one on COVID-19 conspiracy bots on Twitter or this one on influencing political discussion)
   b. There are also engineering blogs like this one from Facebook and this one at Twitter that detail how different platforms are trying to detect bots
   c. We'll gain an understanding of how big the problem actually is, what KPIs measure bot prevalence, and how the solution landscape currently looks

2. **Before April:** Apply research findings in a real data-scraping investigation
   a. We'll use the KPIs we found to analyze an accessible social media platform.
      i. At the moment, we're considering using the Python Reddit API Wrapper to aggregate data from a variety of subreddits.
   b. We'll also gather publicly available user information from the subreddits (account names, account ages, comments & posts) and apply some of the bot detection techniques to this data. We hope to be able to validate some of the claims made by the engineering blogs.

3. **Before Due Date:** Display research synthesis and experiment results in story form on a website.
   a. Craft a narrative from our investigation to display on an interactive website
      i. We're hoping to discover some interesting posts where bots successfully trick users into believing conspiracy theories or propaganda. This would make for a cool storytelling opportunity.
      ii. We'll use modern web frameworks we're familiar with (React, Chart.js) etc.
      iii. We'll also publish our analyses and readings to a GitHub repository that will be referenced heavily in the website.