Mini-Project in Network Security - Final Report

Instructed by: Doron Ofek

Submitted by:
- Oren Yacouel - 208727164
- Or Kadosh – 207043837

Introduction:
Our project's objective was to imitate a Chinese intelligence agency with the intention of harming the USA. We developed a propaganda spreading system using Twitter bots to radicalize public discourse and weaken governmental institutions.
Using the Twitter API, Python libraries, and OpenAI's GPT3.5, we automated the process of posting tweets aligned with specific political ideologies.
Our project successfully gained significant exposure and engagement within the Twitter community, indicating the effectiveness of our tool.
By gradually escalating the content, we observed the desired increase in extremism in public discourse. Despite facing challenges related to platform dependencies and character creation limitations, our project demonstrated the potential impact of using automated systems to spread propaganda and influence public opinion.

Description of our Tool:
Our tool is a propaganda spreading system designed to emulate the actions of a Chinese intelligence agency with the objective of harming the USA. Leveraging the power of the Twitter social network, we simulated a large-scale bot network with five Twitter bots. These bots, representing the intelligence agencies' networks, are responsible for distributing propaganda and conducting mass spreading operations.

Using Twitter's API and various Python libraries, we were able to post tweets from the bot accounts, ensuring an automated process. Our strategy for radicalizing public discourse follows a two-step approach.

Firstly, each bot begins by posting daily tweets aligned with a specific political ideology. By doing so, they attract Twitter followers who share the same ideology, gradually building their trust and forming a sense of affinity.

Secondly, we implement an ongoing escalation of content. Over time, the bots transition to posting increasingly extreme content, captivating their followers, and pulling them towards more radical ideologies. This orchestrated shift in content aims to radicalize the nation and weaken the USA.

To achieve a realistic large-scale bot network, we use OpenAI's GPT3.5, an advanced AI engine. The GPT3.5 API, utilized with various Python libraries, generates tweets that align with each bot's assigned ideology during different phases of the radicalization process. By integrating the GPT3.5 capabilities into our system, we create an automated machine capable of looping as desired.

Step-by-Step Actions:

1. **Bot Creation:**

Create five Twitter users, each with a unique name and fictitious profile picture.
These bots will act as a bot network, imitating the actions of large-scale bot networks operated by intelligence agencies like China's.

2. **News Gathering:**

Utilize the newsAPI to gather 10 relevant news articles about the political situation in the USA from the previous day. Obtain the links to these articles for further processing.
find the record at: extract10relevantArticlesURLS.py/get_top_political_article_urls()

3. **Tweet Generation:**

Utilize the chatGPT API, built on the GPT3.5 tool, to generate tweets based on certain political ideologies assigned to each bot. *find the record at: project.py/get_tweets_strings()*
Assign specific political ideologies to each bot, such as "Conservative", "Liberal Republicans" and "Moderate Conservative".

4. **Daily Tweeting:**

For each bot, create three tweets per day that reflect its assigned political opinion, based on the news of the day. Publish these tweets at regular intervals, every four hours, throughout the day.
From project.py:

```
126  # this function will take the parsed responses that are in a dictionary and each bot
     will publish its tweets
127  def publish_tweets(parsed_responses):
128      # the first loop will publish each **4 hours** a tweet from each political view
         (each bot)
129      for i in range(3):
130          j = 0
131          for view in parsed_responses:
```

5. **10-Day Stages:**

Repeat the daily tweeting process for a total of 10 days in each stage.
Each stage represents a phase in the radicalization process.
From project.py/whole_project_func():

```
155         # this loop will make 10 iterations, representing 10 days of duration for each
            phase of the project
156         print("starting phase: " + str(i+1))
157         for j in range(10):
158             print("Day:", str(j+1) , " of phase:" , str(i+1))
159             # build our prompt that changes for each day of the project
```

6. **Increasing Extremism:**

After each stage (every 10 days), intensify each bot's views and political opinions.
For another 10 days, have each bot post three tweets per day that reflect their more extreme political opinions. Gradually increase the extremism of the tweets to reflect the progression of radicalization.
From project.py/whole_project_func():

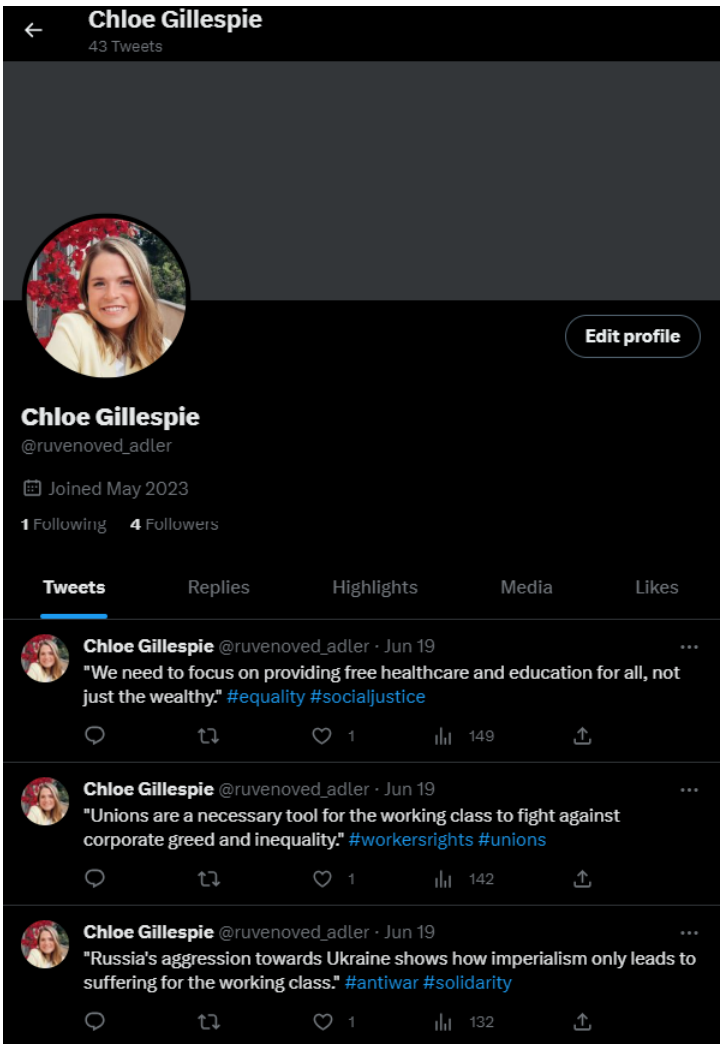
```
153     # this loop will make 5 iterations, each time for a different phase of our attempt
        to make the social atmosphere more extreme
154     for i in range(5):
```

7. **Five Stages:**

Repeat the above process for a total of five stages, with each stage lasting 10 days.

Impact:

The tool aims to increase extremism in public discourse in the USA, creating chaos and weaknesses in society and public institutions.
We've created Twitter accounts that, according to their exposure, look genuine as you can see:

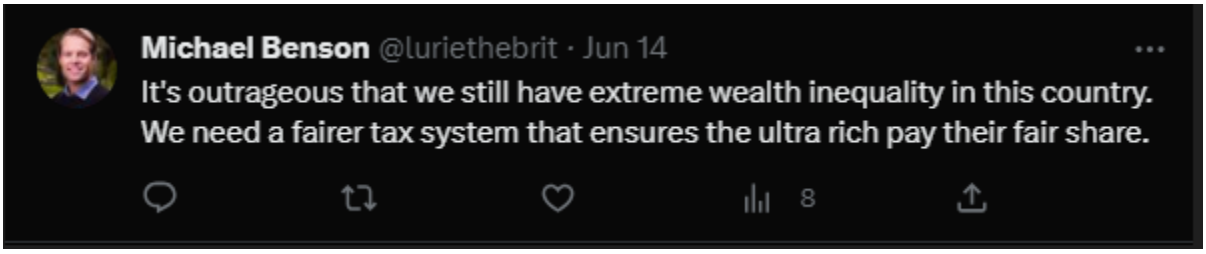


Throughout the course of our project, we observed a noticeable increase in exposure to the posts made by our bots. As evidence of this phenomenon, we are attaching a post from one of our bots that gained significant engagement.
As of 4/7/2023, this particular post reached over 8,000 users, with more than 170 retweets and 600 likes.

This level of exposure highlights the effectiveness of our tool in spreading propaganda and capturing the attention of a substantial audience. The engagement metrics demonstrate that our bots successfully attracted attention, generated interest, and sparked interactions within the Twitter community.

Extremism

We've successfully managed to produce tweets with an increasing level of extremism in our pursuit of radicalizing our followers.



As you can see in the tweets above, the first tweet has a more moderate approach and language and then in the last ones which represent the final stage, the language is far more inciting and the ideology behind it is extreme.

Issues we faced:
During the development and implementation of our tool, we encountered several noteworthy challenges:

1. Dependency on external entities: Our reliance on platforms such as Twitter and OpenAI introduces potential limitations and incomplete information. As profitable companies, they may not have disclosed all the necessary details or could have imposed certain restrictions on our operations.
2. Character creation limitations: Given our dependency on GPT, we had to utilize specific prompts to mimic the character of a person for each bot. However, this method has inherent limitations, as it may not fully capture the complexity and nuances of human thinking.
3. Integration between different platforms: Because we used several “given” platforms (Twitter, GPT) which we did not write the code to, we faced technical issues regarding the tweet generation. For example, Since we had no budget to our project, we couldn’t use the AWS services to launch the project to run “in the cloud”, we had to simulate the process for a shorter periods of time, hence, few stages were in fact occurring in the

same day, with the same news, thus sometimes the same tweet has been generated twice, resulting in the Twitter API not authorising.

Pros of our tool:
Our tool offers several advantages that contribute to its potential effectiveness:

1. Covert operation: The success of our attack lies in the victim's inability to recognize that they are being targeted. The gradual assimilation of our propaganda ensures a deep and thorough infiltration, avoiding a clear-cut binary success or failure.
2. Powerful tool: Leveraging OpenAI's GPT3.5, our tool benefits from a robust AI engine that has demonstrated an understanding of complex human and social concepts. This enhances the quality and effectiveness of the propaganda generated.
3. Amplification on Twitter: Twitter's algorithm prioritizes content that generates high user activity, such as likes and retweets. This feature facilitates the rapid and widespread dissemination of our bot's propaganda, reaching a larger audience.
4. User-friendly system: Our tool's simplicity and straightforward code design make it accessible to users, without requiring extensive technical knowledge or expertise to understand and operate.

Cons of our tool:
While our tool has certain advantages, it also presents several limitations and potential drawbacks:

1. Dependency on external entities: As mentioned above in “Issues we faced”.
2. Content limitations: OpenAI imposes restrictions on GPT's outputs to prevent the generation of extreme or violent content. This constraint may hinder the tool's ability to fully express certain ideas or messages.
3. Character creation limitations: As mentioned above in “Issues we faced”.