Police Bot: Enhancing Social Media Governance with Policing Bots

Milestone 5 Presentation

Group Members:

Students:

- Gabriel Silva
- Cody Manning
- Liam Dumbell
- Nickolas Falco

Faculty Advisor / Project Client:

- Khaled Slhoub

Computer Science Project Instructor:

- Philip Chan

Overview:

- Discussion of Task Completion:
  - Code Improvements
  - Distinguish Algorithms
  - Demo
  - Poster and Ebook Page
- Current Milestone Task Matrix
- Advisor Feedback
- Next Milestone Tasks + Matrix

Code Improvements

- Made the code look more professional.
- Made the code be more articulated for subsequent students who might work on it
- Fixed typos and inconsistencies in variable names
- Made the code more reliable and handled exceptions related to the PRAW library
- Made the code less redundant by removing unnecessary sections and improve efficiency

Distinguish Malicious Bots

Distinguish between the type of bots:

- Scamming Bots (Bad)
  - Accounts sending links or shortened links.
- Harassing Bots (Bad)
  - Accounts sending slurs towards other users.
- Auto Declared Bots (Good)
  - Accounts that warn other users of being a bot.
- List of Known Bots (Good)
  - Accounts that are known to be useful bots.

Distinguish Malicious Bots

```
def further_analysis(bot):
    # checking for shortened links in comments
    # possible fishing bot
    n_links, n_short_links = check_links(bot)
    print(f"{bot.name} has {n_links} links and {n_short_links} short links")

    # checking for profanity
    profanity = count_bad_words(bot)
    print(f"This account cussed {profanity} times")

    # checking opt-out or autodeclared bot
    declared_bot = is_declared_bot(bot)
    print(f"is this acc a autodeclared bot? {declared_bot}")
```

Demo

https://www.youtube.com/watch?v=ZCKQ5wua0G8&ab_channel=JoaoGabriel

EBook Page

[Computer Engineering and Sciences]

Project Name	Framework to Analyze Behavior of Social Media Bots
Team Lead:	Cody Manning
Team Member(s):	Gabriel Silva, Liam Dumbell, Cody Manning, Nickolas Falco
Faculty Advisor(s):	Dr. Khaled A. Shoub, Department of Electrical Engineering and Computer Science, Florida Institute of Technology

****Do not change font size or text color above this message/delete this before completion. The category will be put in by Staff after submission ****

Project Description:
Social media has become a driving force in many people's lives. Connecting with people has never been easier than it is in modern society. Whether it is meeting new people, connecting with old friends, or even showcasing yourself to potential employers; social media is a huge factor in how we socialize with people. This new technology brings new challenges however; as social media has become increasingly infested with entities known as 'bots'. Bots are nonhuman pieces of software that are created for the purposes of automating tasks. A person may create a bot that automatically translates what a user says in Spanish into English, for example. Not all bots are created to help people however. Some people have created bots that serve malicious purposes. These bots act like real people, and may be used to steal information or annoy users who unknowingly interact with them. The rise of AI companions like ChatGPT have made this kind of thing even more prevalent. Our framework is created for the purpose of being able to detect these bots, and possibly differentiate them from the bots that are created for beneficial purposes. The framework was created to work on the 'Reddit' social media platform, but the backbone and ideas of the project could be extended to other social media platforms, with some tweaking depending on the features of the social media it is being adapted to.

Features:
The user will be able to deploy the framework on Reddit using a specific subreddit (which is a collection of topics created by users). The user will also be able to select a specific user (by typing in the suspected users Reddit username). If the user wants to search a specific subreddit, the framework will scan through top posts, newest posts, or the posts that are most popular in a short timeframe. It then asks the user how many posts they would like to search for, and how deep in the posts (how many users) it would like to evaluate. When this is done, it will print out all of the users in the posts it grabbed and give a score based on the likelihood of them being a real human being, or a bot. The framework uses more than one method for detecting bots, and will tell you the results in an easy to understand, color coded result. If the results are unanimous, the framework will say as such, otherwise; it will allow the user to decide at their own discretion. The framework will also give the user insights on whether the given bot is a 'good' bot (one made to help) or a 'bad' bot (one made to harm).

Evaluation:

When designing this framework, accuracy was the key for our measurement of success. It doesn't matter much if the results come quickly if they are wrong. To achieve this, we measure against a master list of known bots (which were scraped from several sources, GitHub and Reddit itself in particular). We were shooting for about an 80% accuracy method in detecting whether a user was a bot. When using our known bot list, and a list of known real human accounts, the accuracy rating was well within our desired output. Our timing desire was no more than about 10 seconds per account lookup, and this was unfortunately not really feasible within the context of how the program functioned. It really came down to speed or accuracy, and the team decided that accuracy was what we wanted to focus on.

Major Challenges:

There were a lot of challenges we encountered and overcame during the process of this project. Particularly detecting bots and distinguishing the good from the bad bots. This is still a widely researched topic in Computer Science, so we were working blind for a lot of this. We found that one detection algorithm alone was simply not sufficient for proper detection accuracy, and implemented a second detection algorithm to supplement the first. Realistically, the more detection algorithms that get added the better. For the future of this project, we should get as many as we can. The second and biggest challenge we faced was detecting the 'nature' of the bots we found. There is no real clear answer on this, so we had to work with the information we were given. One notable observation was the inclusion of links. We couldn't find a good reason for bots to direct you outside of the Reddit platform, so it was immediately flagged as suspicious if they wanted you to leave the site and go somewhere else (especially if the outside link was obscured with a link shortener). There is no perfect science for this project, so it is something that needs to be built on more in the future.

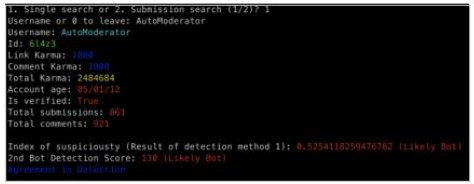


Figure 1: Searching a known bot by username

Poster

NORTHROP GRUMMAN
ENGINEERING & SCIENCE
STUDENT DESIGN SHOWCASE
FLORIDA TECH

Enhancing Social Media Governance with Policing Bots

Cody Manning, J. Gabriel Silva, Liam Dumbell, Nickolas Falco
Faculty Advisor: Khaled Slhoub, College of Engineering and Science - Electrical Engineering and Computer Science, Florida Institute of Technology

Project Description

Social media has transformed how we connect, from meeting new people to showcasing ourselves to potential employers. However, it's also plagued by bots—automated software designed for various purposes, including malicious activities. These bots mimic human behavior, posing risks like data theft or annoyance to users. AI tools like ChatGPT have exacerbated this issue even further. Our framework has been developed to detect and differentiate between beneficial and malicious bots on Reddit. Expansion to other social media platforms is feasible.

Evaluation

Accuracy was paramount in designing our framework. We aimed for 80% accuracy in detecting bots, using a master list from various sources. While our desired timing was under 10 seconds per account lookup, achieving speed and accuracy was challenging. Ultimately, prioritizing accuracy over speed was our team's decision.

Challenges

We faced challenges in detecting and categorizing bots, necessitating multiple detection algorithms. Determining bot nature was also challenging, we flagged bots redirecting outside Reddit as suspicious as one of our methods. The project lacks a perfect solution, highlighting the need for further research and algorithm development in the future.

Features

The framework enables users to deploy a Police Bot on Reddit within a chosen subreddit and select a specific user by their Reddit username. It scans through top, newest, or popular posts in a subreddit and asks users to specify the number of posts and depth of evaluation. After scanning, it provides a list of users with scores indicating the likelihood of being a bot or human, using multiple detection methods. Results are presented in a clear, color-coded format. If the results are unanimous a message saying so will be displayed; otherwise, the framework will indicate there was a disagreement. Additionally, the framework offers insights on whether identified bots are "good" or "bad" based on their purpose.

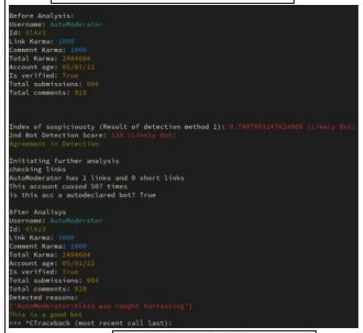
Key Features:

- Deployable on Reddit within a chosen subreddit.
- User selection by Reddit username.
- Scan options: top, newest, or popular posts.
- Specify number of posts and depth of evaluation.
- Multiple bot detection methods.
- Clear, color-coded result presentation.
- Confirmation of detection results.
- Bot classification: "good" or "bad."

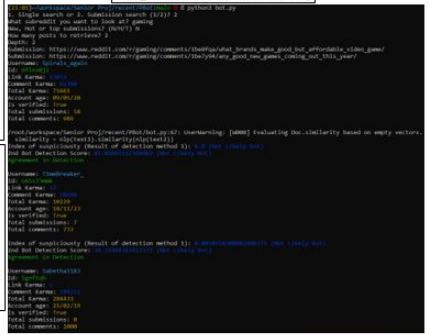
Project Expansion

Our bot detection framework is versatile, easily transferable to other social media platforms like Twitter, Instagram, and Facebook. By adjusting API requests, account data parameters, and other specifics, it can seamlessly adapt to the unique features and requirements of each platform, ensuring effective bot detection across various social media channels.

Submission Search (Subreddit)



Single Search (Username)



Current Milestone Task Matrix

Task	Completion	Cody	Gabriel	Liam	Falco	To Do
Find and implement more detection algorithms	80%	20%	20%	10%	30%	Finalize our final detection algorithms.
Figure out the distinguishing module	70%	20%	30%	20%	20%	Refine the methods we use to define maliciousness so that we can get a clearer picture on a bots nature.
Ebook page and Poster	100%	50%	0%	50%	0%	

Advisor Feedback

- Satisfied with our current progress.
- Suggested changes on the poster to be less verbose and with more images.
- Suggested that we continue improving the maliciousness detection and find different ways to distinguish bots.
- Guided us towards milestone 6 with ideas for the decide method to report bots to reddit.

Milestone 6 Plan

- Finish the Bot Detection Algorithms
- Finish the Bot Distinguishing Algorithms
- Organize the code in a way that makes it simple to add features
- Work on organizing the repository for subsequent students
- Update the options menu to implement the Inquirer library
- Create a user / developer manual and requirements list for framework
- Finalize Framework (Database, Algorithm tuning, command line UI, Final Pushes to GitHub)
- Create Final Demo to test entire framework

Next Milestone Task Matrix

Task	Cody	Gabriel	Liam	Falco
Finalize the detection algorithms	20%	10%	10%	60%
Finalize the maliciousness algorithms	10%	50%	20%	20%
Test the framework as a whole	30%	10%	50%	10%
Create developer / User Manual	70%	10%	10%	10%
Final Demo	25%	25%	25%	25%

This concludes our presentation, Thank You