Police Bot:

Enhancing Social Media Governance with Policing Bots

Milestone 6 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
 - Testing Metrics
 - Decide Module
 - Demo/Commercial
 - Update on Poster and Ebook Page
- Milestone Completion Task Matrix
- Advisor Feedback
- Lessons Learned

Testing Method

Using List of Known Humans:

- Compiled using the moderator lists of the top subreddits (where bot moderators are declared as bots, eg. AutoModerator)
- Used to test the percentage of false positives (Humans identified as Bots)
 List of Known Bots

Using List of Known Bots:

- Compiled using various public list of Reddit Bot accounts
- Used to test the percentage of false negatives (Bots identified as Humans)

Testing Metrics

Positive Assessments (Known Bot List)

- Objective: 70%-80% correct bot detection
- Results: 320/424 (bots) - - 76% Bots identified as Bots

Negative Assessments (Known Human List)

- Objective: <20% false positives
- Results: 102/651 (humans) - 15% of Humans identified as Bots

Code Improvements

- Enhanced organisation
- Enhanced maintainability
- Enhanced reusability
- Enhanced Scalability

Decide Module

Decide to report or not a scanned bot:

- Scamming Bots or Harassing Bots (Bad)
 - Are provided with a link to the report page and a tutorial.
- Auto Declared Bots or Known Bots (Good)
 - Ask the user whether to offer the option to report good bots.

Decide Module

 \circ

```
1 decide.py
  1 from colors import cyan, reset
 3 link = "https://www.reddit.com/user/"
 4 def decide(account, ignore_list):
        if 'exiting' in ignore_list:
            print(f"url: {cyan}{link}{account.name}{reset}")
            return True
        if 'all' in ignore_list or account.good_bot:
            return False
        if account.good_bot == '0' and not 'good' in ignore_list:
            print(f"Autodeclared bot usually not harmful.")
            x = input("Want to report? (y?) (0 to ignore all)")
 12
            if x.rstrip().lower() == 'v':
                print(f"url: {cyan}{link}{account.name}{reset}")
14
                return True
        else:
            if len(account.reasons) > 0:
                print(f"url: {cyan}{link}{account.name}{reset}")
                return True
            elif not 'inconclusive' in ignore_list:
                print(f"No decisive reason to report this bot.")
                z = input("Want to report? (y?)")
                if z.rstrip().lower() == 'y':
                    print(f"url: {cyan}{link}{account.name}{reset}")
        return False
```

Demo/Commercial

```
human: Constant-Anybody-783
                                                           Alledon welves polacita
good bot: AutoModerator
                                                          his mount has seed professing will kines
bad bot: Angelakayla5
                                                           With terrount is authoritized total
                                                           proper or 6 to Inner Deptimality
                                                           MATERIAL PROPERTY.
                                                           STATE SHEET SHARE
                                                           nerships Than
                                                          TITL SPEAKING PARTY TRANS
                                                          on that Speculate Spread
                                                          nd Bed Debruktion Score: I
                                                          Ministry Author assistate
                                                          Operations for Witches and P (Revision) Educa-
                                                          his amount has sent profunity it have:
                                                           With married a partiagnal great best F
                                                         MINISTER THACK ARTERUATED " SCOUNT FOR LAWYOR PRESENTANTED SCOUNT DAY LOTTE THA JURISTINSTON
                                                           ment men "Ang Lating Se" motore fina pictore a part mit mit better in a court. Dis incide 221 statement 2 to 6
                                                         District And Secretary School Store
          - Mar D - # < O + F # 5 K + + 2 K + 6
                                                                                                                                   Better - Date to - TT &
```

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. Sihoub. Department of Electrical Engineering and Computer Science. Florida Institute of Technology			

**Do not change font size or text color above thi category will be put in by Staff after submission *

Project Description:

Social mode has become a driving force in many posbut in it is in mode mode, without in it is most showcasing yourself to potential employers; social methous and pose of the position of the position of the social mode and the position of the position of the social mode and the position of the position of the Method that we contain place of critical position of the mode position of the position of the position of the mode position of the position of the position of the mode position of the position of the position of the mode position of the social media platforms, with some breaking depending on the position of the position of

Features:

The user will be able to deploy the framework on Regide topics created by users]. The user will all so be able to be Regide users growing the state of the specific processors and the specific memors post, or to the posts that are most popular in a the newset post, or the posts that are most popular in as they would like to search for, and how deep in the post is done. It will print out all of the users in the post it is post in the post in a real human being, or a bot. The framework us you the results in an easy to understand, color coded as ya as such, otherwise; it will allow the user to decide user insights on whether the glow not be it a 'good' but to

valuation:

When designing this framework, accuracy was the key if the results come quickly if they are word;. To achie (which were scraped from several sources, Githbia an 80% accuracy method in detecting whether a user known real human accounts, the accuracy rating was more than about 10 seconds per account lookup, are control of how the program functioned. It really came accuracy was what we wanteed to floxus on.

... ...

here were a bet of fullenges we encountered and overame during the process of this project. Particularly detecting bets and distinguishing the good from the bad bets. This is still a widely researched topic in Computer Science, so we were working Bird for a bit of this. We found that one detection algorithm above simply not sufficient for proper detection auxaxive, an implemented a second detection algorithm to supplement the first. Bealstically, the more detection algorithms that get added to be better. For the future of the project, we doubt and a many are on. The second and begate cludinger were found visible to the project, we doubt and a many are on this, so we had to work with the information were agen. One extends allowershown such inclination of first. We could find algorithms for both one of the state of



Figure 1: Searching a known bot by username

[Computer Engineering and Sciences]

Project Name Framework to Analyze Behavior of Social Media Bots

Team Lead: Cody Mannin

Team Member(s): Gabriel Silva, Liam Dumbell, Cody Manning, Nickolas Falco
Faculty Advisor(s): Dr. Khaled A. Slhoub, 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. Some people have created bots that seve malicious purposes. These bots act like real people, and may be used to stal information or annow users who unknowingly interact with them. 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 'Reddi'r 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 will also give the user insights on whether the given bot is a "good" bot (one made to help) or "abd" 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 tall a widely researched topic in Computer Science, so we were working blind for a lot of this. 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.



Poster



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 annovance to users. Al tools like ChatGPT have exacerbated this issue even further. Our framework ha 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 outside Reddit as suspicious as one of our methods. The project lacks a perfect solution, highlighting the need for further research and algorithm velopment 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 nurnose

Key Features:

Deployable on Reddit within a chosen

- subreddit Ilser 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 challenging, we flagged bots redirecting 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)





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. Al 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.

- 80% detection accuracy goal
- Targeted lookup time under 10s
- Prioritized accuracy over speed
- Challenges Multiple detection algorithms required
- to meet goals Determining bot nature (Malicious /
- Helpful) · Future research and algorithm
- development needed

Functions of Police Bot Framework:

Detection

- · Text Frequency Analysis
- · Account Data Analysis (age, posting intervals, verification status, etc)

Distinguish

Checking spam, use of profanity, redirect links, etc.

Decision

Made based on the severity of malicious activity

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."

Submission Search (Subreddit)

Single search or 2. Submission search (1/2)

ter Analysis ername: AutoModerator

sk Karma: count age: 05/01/12 verified: True

Project Expansion

- Easily transferable to Twitter, Instagram, and Facebook
- Adjust API requests and account data
- parameters to do so Ensures effective bot detection across social

media platforms without having to rewrite detection methods



Milestone Completion Task Matrix

Task	Cody	Gabriel	Liam	Falco	To Do
Finalize the detection algorithms	15%	10%	50%	25%	N/A
Create the decide module	10%	40%	30%	20%	N/A
Finalize the maliciousne ss algorithms	5%	40%	20%	5%	N/A
Test the framework as a whole	40%	40%	10%	10%	N/A
Create developer / User Manual	70%	10%	20%	0%	N/A
Final Demo	10%	50%	15%	25%	N/A

Advisor Feedback

- Satisfied with our current progress.
 - As aforementioned in early presentations the project's goal was to build from the ground up a framework that polices a social media looking for bots
- Invited us to continue working on this project even after we are done with the class.
- Mentioned writing and publishing a paper with our findings,
- Or making the project available to students in the future.

Lessons Learned

This project has been quite an ordeal, as aforementioned in early presentations the project's goal was to build from the base up a framework that can be improved in the future.

- Have a clear plan, and stick to it
- Understand API and Platform Limitations
- Start early
- Test early

This concludes our presentation, Thank You