Project title: Enhancing Social Media Governance with Policing Bots

Team Members:

- Cody Manning cmanning2020@my.fit.edu
- João Gabriel Silva jsilva2021@my.fit.edu
- Liam Dumbell ldumbell2021@my.fit.edu
- Nickolas Falco nfalco2021@my.fit.edu

Advisor and Client:

- Khaled Slhoub kslhoub@fit.edu
- Affiliation: Florida Institute of Technology

Date(s) of Meeting(s) with the Client for developing this Plan:

- First Meeting - January 19, 2024
- Future Meeting Frequency - Bi-weekly (swapping to weekly and twice weekly as needed)

Goal and motivation:

Our framework is made with the intention of detecting “bots” on the social media platform reddit.com. The website is having a serious problem with bots, as they attempt to maliciously take information from users, or trick them into giving away vital information about their accounts or personal lives. With our framework, we hope to be able to firstly detect the bots, then determine the purpose of the bot (some bots are not malicious in nature, and are in fact beneficial to the users), and finally decide whether the bot should stay on the platform or not. In most cases, the good bots should stay on the platform, and malicious bots should be removed from the platform. This framework will lead to a better user experience for reddit, and can hopefully be extended to work on any other social media platform, in an attempt to keep users safe and engaged with their platforms of choice.

Approach:

Key Features:

- Detect Bots
  - Our first feature would allow the client to detect artificial users (bots) in a social media platform. Our first feature would revolutionize the way clients interact with social media platforms by equipping them with a tool designed to detect artificial users, commonly referred to as "bots." By analyzing an account's activity we will detect suspicious behavior that would indicate whether or not the account is run by a bot.
- Distinguish
  - Our second feature will be the ability to distinguish beneficial bots from malicious ones.

    We will need to have more discussions on the ‘ranking’ of the severity of the types of bots. A non-harmful bot is one that is determined to have no negative effects on the social media platform in which it is active. A harmful bot is one that is determined to have some form of negative effect on the social media platform it is active on. This can range from minor negative effects such as spam on a user’s post or profile, to majorly negative effects such as scam bots or bots that

    - maliciously influence political opinions of the social media's user base on a large scale. By assessing factors such as the bot's origin, its interaction patterns, and its impact on user experience, we can categorize bots into two distinct classes: non-harmful and malicious.
- Decide
    - Our third feature will be deciding what to do with the account after determining that it is a bot and its level of maliciousness. If we find that the bot account is breaking the terms of service of the platform it is active on, we plan on determining the level of maliciousness, and then determining whether reporting the account to the platform administrators is necessary. If the rule broken is determined to be relatively harmless, the tool will flag the account to collect data from it to improve itself in the future.

Algorithms and tools (libraries/api/frameworks/languages) for the key features:

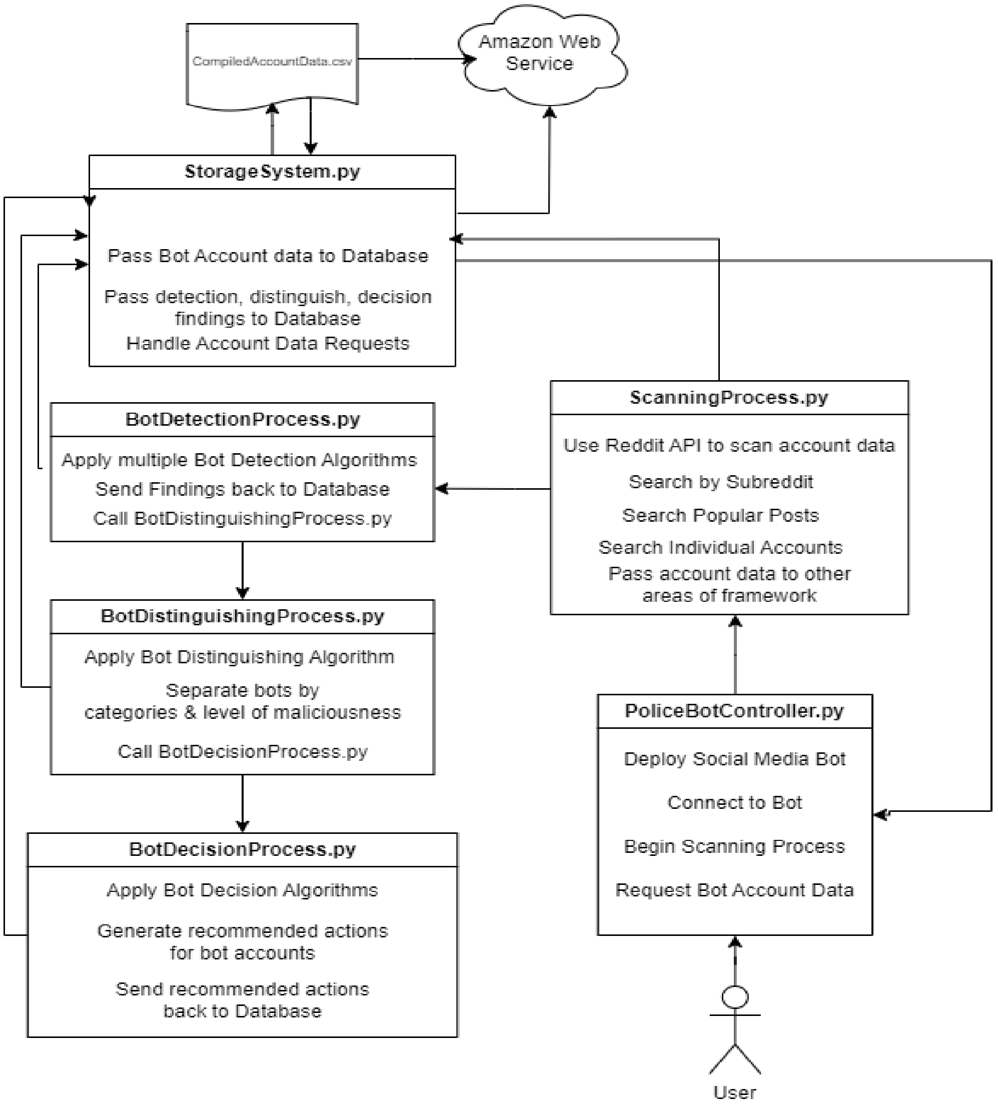
- Reddit API:
    - Using the Reddit API allows us to pull data from the social media website to perform analysis on. This is useful for bot detection, distinguishing and decisions.
- Python:
    - Allows us to automate the process of discovering, scanning, downloading, organizing and analyzing Reddit account data. This is useful for bot detection, distinguishing and decisions.
- PRAW Python Library:
    - Allows us to directly interface with the Reddit API which we use to automate sending API requests and receive Reddit account information. This allows us to perform analysis on the account data to determine if the account is run by a bot, distinguish known bots into different categories and levels of maliciousness, and decide on a recommended action to be taken against the account.
- CSV Python Library:
    - Allows us to create CSV files filled with Reddit Bot Account information that aids in distinguishing bots into categories by allowing us to analyze trends in their account data.
- MySQL Framework:
    - Allows us to store data sent to intermediary CSV files on a MySQL database that can be hosted locally or using AWS. This aids in distinguishing bots into categories by allowing us to analyze trends in their account data on a large scale which will be crucial when we finalize the findings of our project.

Technical Challenges:

- Group members need to continue developing their experience using virtual environments for Reddit and using the Reddit API.
- Group members need to continue developing their skills creating bots with the proposed functionality. To achieve this, additional research into Bot Distinguishing and Decision methods must be done.
- Group members need to continue developing their skills using the libraries and other methods for developing bots for social media platforms.

- Group members need to continue developing their HTML skills to properly understand and use the different social media APIs.

Design: System Architecture Diagram


CompiledAccountData.csv
Amazon Web Service
StorageSystem.py
Pass Bot Account data to Database
Pass detection, distinguish, decision findings to Database
Handle Account Data Requests
BotDetectionProcess.py
Apply multiple Bot Detection Algorithms
Send Findings back to Database
Call BotDistinguishingProcess.py
ScanningProcess.py
Use Reddit API to scan account data
Search by Subreddit
Search Popular Posts
Search Individual Accounts
Pass account data to other areas of framework
BotDistinguishingProcess.py
Apply Bot Distinguishing Algorithm
Separate bots by categories & level of maliciousness
Call BotDecisionProcess.py
PoliceBotController.py
Deploy Social Media Bot
Connect to Bot
Begin Scanning Process
Request Bot Account Data
BotDecisionProcess.py
Apply Bot Decision Algorithms
Generate recommended actions for bot accounts
Send recommended actions back to Database
User

Evaluation:
The success of the project will be evaluated based on the following criteria:

- Speed: Since the project relies heavily on web scraping and API calls, the project can get a little bit slow when lots of data is being requested. We are hoping to make it never take longer than 30 seconds to receive data, but this is still to be worked on.
- Accuracy: The accuracy of the project is the most important feature, our goal is 80% accuracy at correctly detecting bots.
- Reliability: Reliability is the second more important measure of success. Realistically, the program should return with the same result every single time (assuming that the data we are given about the user remains the same).

Progress Summary:

Module / Feature	Completion %	To do
Detection Module	70%	Refine the algorithm for accuracy, begin adding more algorithms that work in tandem with the already established algorithm.
Distinguish Module	0%	Begin work on detecting malicious or beneficial bots.
Decide Module	0%	Create a decide module that figures out what to do with detected bots.
Backend Database	80%	Perhaps begin implementing AWS as our database of choice.

Milestones for the second semester:

- Milestone 4 (Feb 19):
  - Figure out the next algorithm we want to use to work with the established detection algorithm we already have.
  - Begin research on how we are going to implement the decide module
  - Clean up and decide what we are going to do with the database
  - Try to speed up the current working algorithm for the detection module
- Milestone 5 (Mar 18):
  - Work on the deciding algorithm
  - Test accuracy and speed of the two modules, particularly the detection algorithm
  - Combine the two modules into a single framework environment
  - Create poster and ebook page for Senior Design Showcase
- Milestone 6 (Apr 15):
  - Work on and finish the reporting module
  - Clean up and merge all of the modules together
  - Test and bugfix
  - Demo the framework
  - Create Demo Video

Task matrix for Milestone 4:

Task	Cody	Liam	Gabriel	Falco
Research detection algorithm to work in tandem with the current one, and implement it if possible.	Research	Research	Research	Research
Work on efficiency for the current detection module	Work on the data structures	See if the API calls can be sped up	Research and implement algorithms for efficiency and accuracy	Assist where needed in efficiency update
Work on the database functionality	Figure out AWS	Implement functionality in the database module	Integrate this module to the main program	See if the way the database is used can be improved
Research methods for the deciding module	Research	Research	Research	Research

Description (at least a few sentences) of each planned task for Milestone 4:

- Task 1 and 2: While our current detection method technically works, it has a lot of room for improvement. As we mentioned in earlier tasks, no single bot detection algorithm is good enough to stand on its own. We hope to find another algorithm to work in tandem with the current detection algorithm we use. This should improve accuracy greatly. We also need to see if we can speed up the process with proper data structures, but it is unknown if the web scraping bottleneck can be overcome with proper programming, or if that is something that is simply a hardware restriction.
- Task 3: The database module is what stores all of the data we request. We hope to complete this module by the end of this milestone. Hopefully fully integrating this to the main program will help speed up the framework and solve some of the problems that exist in the previous milestone. We think once the framework has the data collected, every subsequent run (on the same data) will benefit from greatly increased speeds.
- Task 4: We need to begin work on the second stage of the project, which is the deciding module. We need to figure out how exactly we plan to separate malicious bots from beneficial bots. There are lists of beneficial bots on reddit, and perhaps we can use this as a baseline for the work we plan to do. The problem will be in the malicious bots, as the definition of maliciousness will vary from user to user. There may be wordlists on the

internet of common scam tactics, and we will be looking through these for examples of what to look for.

Approval from Faculty Advisor

- "I have discussed with the team and approve this project plan. I will evaluate the progress and assign a grade for each of the three milestones."
- Signature: ________________________________ Date: ________