

Enhancing Social Media Governance with Policing Bots - Milestone 2 Evaluation

Authors:

- Cody Manning - cmanning2020@my.fit.edu
- Gabriel Silva - jsilva2021@my.fit.edu
- Liam Dumbell - ldumbell2020@my.fit.edu

Client:

- Dr. Khaled Slhoub - kslhoub@fit.edu

Progress of current milestone (Task Matrix):

| Task | Completion | Cody | Gabriel | Liam | To Do |
|--|------------|------|---------|------|--|
| Research as many social media APIs as possible (with the possibility of switching from twitter if it becomes unfeasible) | 100% | 33% | 33% | 33% | |
| Gain a rudimentary understanding of the API and environment of whatever new social media platform we choose | 60% | 20% | 20% | 20% | Keep expanding knowledge of our new social media platform of choice. |
| Develop a system to collect basic data on social | 100% | 25% | 50% | 25% | |

| | | | | | |
|--|-----|-----|-----|-----|---|
| media accounts | | | | | |
| Research known bot detection methods | 40% | 20% | 10% | 10% | This is an ongoing process in our project, we need to break down and choose a detection method as our jumping off point |
| Research and potentially find a way to store the data we collect | 50% | 10% | 10% | 30% | For the moment, we just store the data in a standard csv file, a DB would be more efficient for larger datasets, so we should consider it |

Discussion for each accomplished task for the milestone:

- Task 1: With the realization that Twitter was unfeasible because of recent API changes to the platform, we turned our attention to other social media websites. YouTube was considered, as was FaceBook and Instagram, but Reddit was the one we decided to go with, as it has a bunch of bots; both harmful and beneficial. Having these known bots will make it easier for us to have a jumping off point for our bot detection.
- Task 2: We did a lot of research into the various API's available for each social media platform, with Reddit being the one we landed on. This is because, even though there are restrictions on what can be done with the API without paying, it won't matter for our case since it is so small in scope. So we have done work finding tools to communicate with the API and use it to work on our project. We ended up finding the praw (python reddit api wrapper) library, and are currently using it as the backbone of our tool.
- Task 3: We have created a basic tool to check through subreddits, and collect data on the users who are commenting on the posts. Currently, the program asks

for a subreddit, and allows the user to pick the depth (which is how deep the bot parses through comments), the type of post (current options are Hot, New, and Top posts), and the amount of posts you want to scan. This program scans, collects the data on the users, then prints it to the terminal and saves it to a .csv file. It runs a little slow right now, but we're not sure if it is a code issue or an api issue, we think it is API related.

- Task 4: We have done cursory reading of the papers we have been given from Dr. Should, but there is still more to be done here. This task is going slower than we would like, but we have begun doing research on known bot detection methods. Our advisor/client has given us a large list of academic papers on this subject, and we have done brief readings on it. This is definitely something we need to focus on more heavily for the following milestone.
- Task 5: We wanted any method of storage to start with, and we have gone the simple route and just saved it to a local csv file to start with. This is okay for small datasets, but on larger datasets it's just going to become messy and slow. We chose this because it was quick and simple, but for the end product a better way will have to be implemented.

Discussion of contribution for the current milestone:

- Cody Manning: Cody shared the task for research, along with assisting in the demo. He was responsible for researching social media sites, ultimately making the decision for the group to switch to Reddit as the social media platform of choice. Cody and Gabriel decided on the 'praw' library for python Reddit API functionality. Cody assisted Gabriel in writing the demo. Cody also wrote this document.
- Gabriel Silva: Gabriel was the main person responsible for the demo programming. He helped research bot detection methods, Gabriel found the 'praw' library for the API wrapper. This of course, also shows that he was assisting in the social media research and API research. Gabriel probably understands the API the best of the three of us, so he will also be helping educate Cody and Liam in the proper ways of using it.
- Liam Dumbell: Liam has the most experience with databases, so he is doing the brunt of the storing work. Although we currently don't have any kind of database implemented, Liam is working towards potentially fixing the problems outlined in our task 5 mentioned above. Liam also researched social media sites, bot detection algorithms, and reddit API methods.

Plan for the next milestone (Task Matrix):

| Task | Cody | Gabriel | Liam |
|--|------|---------|------|
| Improve data collection system | 50% | 25% | 25% |
| Research and decide on a single starting bot detection method | 33% | 33% | 33% |
| Start implementation of chosen bot detection method | 25% | 25% | 50% |
| Create a working demo of rote bot detection (with the data collection integration) | 25% | 50% | 25% |

Discussion for each task of the next milestone:

- Task 1: The way we collect data is a little slow, we need to see if this is simply an API problem, or a problem with the way we collect data. If it's the API, there isn't much we can do about it, but if it is our program's fault, we need to improve it. We also need to see if anything needs to be changed to implement the rest of the features (in this case, a simple detection algorithm).
- Task 2: Our advisor has told us to continue researching bot detection methods, as there isn't really one best solution yet. We are hoping to lock one in for this milestone, so we can finally get our project off the ground.
- Task 3: Once we decide on a system, we need to get a simple version of it coded, we can get it working with the data collection program.
- Task 4: This is where the previous three tasks will come together. We are hoping to have the data collection and the simple bot detection working in tandem at this step.

Dates of meeting with Client:

- October 27th, 2023

Client Feedback on Milestone 2:

- See faculty feedback below

Dates of meeting with Faculty Advisor:

- October 27, 2023

Faculty Advisor Feedback on Milestone 2:

- Task 1: Dr. Shloub was happy that we were able to rebound easily to a new social media platform, telling us that Reddit seemed like a good place to begin the project. He did remind us that the program would someday be used for any social media platform, not necessarily just Reddit, but was happy we found one that will work for our use cases.
- Task 2 and 3: We showed Dr. Shloub our demo, and he was happy with the progress on our knowledge of the API. He liked the demo, and had a few questions on what certain areas of the demo were (the depth in particular). He told us to continue on with what we were doing.
- Task 4: Dr. Shloub told us to continue researching bot detection methods, using the documents he sent. He also brought another paper to our attention that he thought would be useful for our efforts. He reminded us that we should try to lock in one type of detection algorithm, rather than just trying to implement them all.
- Task 5: Dr. Shloub mentioned that a proper way of storing data was important, but not to spend too much time researching this when our efforts are better served elsewhere.

Faculty Advisor Signature _____ Date: _____

Evaluation by Faculty Advisor:

- Faculty Advisor: detach and return this page to Dr. Chan (HC 214) or email the scores to pkc@cs.fit.edu
- Score (0-10) for each member: circle a score (or circle two adjacent scores for .25 or write down a real number between 0 and 10)

| | | | | | | | | | | | | | | | | |
|---------------|---|---|---|---|---|---|-----|---|-----|---|-----|---|-----|---|-----|----|
| Cody Manning | 0 | 1 | 2 | 3 | 4 | 5 | 5.5 | 6 | 6.5 | 7 | 7.5 | 8 | 8.5 | 9 | 9.5 | 10 |
| Gabriel Silva | 0 | 1 | 2 | 3 | 4 | 5 | 5.5 | 6 | 6.5 | 7 | 7.5 | 8 | 8.5 | 9 | 9.5 | 10 |
| Liam Dumbell | 0 | 1 | 2 | 3 | 4 | 5 | 5.5 | 6 | 6.5 | 7 | 7.5 | 8 | 8.5 | 9 | 9.5 | 10 |

Faculty Advisor Signature _____ Date: _____