User Strategies for Social Bot Detection

Simge Tekin
stekin@umd.edu
University of Maryland
College Park, MD, USA

Patrick Gough
pgough@umd.edu
University of Maryland
College Park, MD, USA

Anastasios Toumazatos
tasos@umd.edu
University of Maryland
College Park, MD, USA

KEYWORDS

Social bots

ACM Reference Format:
Simge Tekin, Patrick Gough, and Anastasios Toumazatos. 2024. User Strategies for Social Bot Detection . In *CMSC732*. ACM, New York, NY, USA, 13 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

As social media becomes the primary medium for staying updated on current events, spreading opinions, and acquiring information, the presence of 'social bots' has become increasingly prevalent. We define social bots as automated agents that operate on social media platforms, generating content and interacting with human users in a way that mimics human behavior, in accordance with prior studies in the field [7] [1]. These bots are deployed on a large scale to engage in various activities ranging from benign actions such as entertainment to malicious ones, like creating fake public opinions, manipulating users toward certain ideas, and spreading misinformation. According to a 2020 study, bots are estimated to comprise 15% of Twitter's (currently known as X) user population [14]. While bots have recently drawn the public's attention due to their alleged influence on recent political events [15], and prevention methods suggested by the platform owners [8], they have been present since the early days of social media. Researchers have engaged in a continuous "cat and mouse" game with bots since the 2010s, beginning with supervised machine learning techniques for detection [3, 22]. Over the past decade, many automatic detection techniques have been proposed, including neural networks [2, 10] and more recent methods that apply large language models (LLMs)[6]. Additionally, various tools are made publicly available for practical use [5, 21].

However, as more advanced techniques emerge and get adopted for bot detection, the same methods are also contributing to the emergence of more advanced bots that are more similar to legitimate users [4, 20]. As bots evolve rapidly, detection methods are not always deployed quickly enough to clean out the social media platforms, resulting in real users frequently encountering bots in their social media feeds. Understanding how users adapt to new types of bots is essential for developers and researchers to create new techniques to stay ahead of bot developers. Moreover, these insights can be used to inform the development of better education campaigns for users. Based on the intuition about user adoption of automatic detection tools, future tools can be refined to be more user-friendly and effective. In the light of these, we investigate the following research question:

RQ1: *What strategies do non-expert users adopt to detect social bots in their social media feeds?* with the sub questions investigating:

- *Variations of users' strategies and confidence levels under different communication scenarios*
- *Users' openness towards the automatic detection tools and their expectations from such tools*

We conducted five semi structured in-person interviews, and provide the following qualitative insights:

- We identify specific visual and linguistic cues users associate with bots, such as blurry profile photos, basic usernames, and unnatural language patterns in posts and replies. We also record traits that participants believe make bots harder to detect, such emotional language and adaptive responses.
- We observe that users hold varying definitions of social bots, ranging from more technical, like *'AI users'* to more practical, like *'fake profiles'* and *'scamers'*. This lack of uniformity could have a statistically significant effect on their detection strategies in a larger scale study.
- We find that, while none of the participants is actively using any publicly available tools to detect bot-generated content in social media, all were open to adopting one in the future, with the two key prerequisites being the *ease of use* and *free access.*
- We find that users' confidence in bot detection is higher in active engagement scenarios compared to passive ones, driven by the ability to elicit direct responses.

2 RELATED WORK

Social media platforms have become major venues for information exchange; however, the increasing presence of bots adds complexity to the user experience while exerting social influence on users. Previous work highlights the influence and evolution of bots in parallel to the emergence of advanced detection methods.

Bot Behavior and Influence. Bots vary widely in purpose, as shown in the work of Stieglitz et al. [16]. Their research distinguishes between benign bots, like those aggregating news, and adversarial bots used in political influence or spam, offering a clearer view of the context in which everyday users must make judgments about suspicious accounts on social media. The influence of social bots on public discourse has been investigated in studies addressing misinformation spread. Shao et al. [15] analyzed the roles bots

CMSC 732, December 10, 2024, College Park, MD

https://doi.org/XXXXXXX.XXXXXXX

played in amplifying low-credibility content on Twitter during the 2016 U.S. election. By amplifying false information early in the news cycle, bots were able to reach broad audiences, exploiting the trust of human users. This study provides insights into the effect of bot-driven misinformation, highlighting the importance of assessing the legitimacy of online accounts. Yang and Menczer [20] uncovered botnets using AI tools like ChatGPT to generate convincing interactions, complicating detection. Similarly, Yardi et al. [22] showed how bots exploit platform features to blend in with legitimate users, adapting to trends and user behaviors.

Early Detection Methods. Primary bot detection approaches have leveraged traditional machine learning techniques. Chu et al. proposed a method that uses Twitter messages and user metrics calculated on account metadata to determine if a posting account is a bot, incorporating a Bayesian classifier [3]. While this study demonstrated that it is possible to achieve strong accuracy, it did not discuss incorporation into usable tools. Davis et al. created a much simpler classifier that comes with a public website called BotOrNot [5], which leveraged the Twitter API to query information about a suspected account. Using thousands of features, including geographic location, timings of content generation, and sentiment analysis, their model, incorporating a random forest classifier, reached 95% accuracy. However, its reliance on Twitter API access led to its discontinuation in 2023, highlighting the need for bot detection methods that do not depend on third-party APIs. This research produced a user-friendly tool, although, to our knowledge, there has been no further research evaluating its utility. Nasim et al. recruited three expert Twitter users to hand-label Twitter accounts as users and bots and used it as a testing dataset for their classifier [11]. The users reported difficulties with labeling accounts based solely on the posts, such as the difference between a human and a bot account being "not at all clear" and that some accounts are "part automatic and potentially part human." This study compared the accounts that users flagged as bots to the output of BotOrNot, determining that BotOrNot predicted 65% of the identified bots were actually bots. We build upon this research by asking non-expert social media users about their strategies in everyday use, intending to learn more about practical approaches for detecting bots.

Advanced Detection Methods. Deep neural networks have significantly improved bot detection. Kudugunta et al. [10] introduced a deep neural-based approach adopting LTSM architecture for both tweet and account-level detection, yielding high accuracy. Wu et al. [19] combined active learning with neural models, yielding high performance on their dataset. The advance of LLMs has introduced new challenges, creating highly human-like bots that evade traditional detection tools. Tang et al. [17] emphasized the necessity of developing robust methods that can detect sophisticated LLM-generated text in their study, which explores black-box and white-box methods for detecting generated text. Jiang et al. [9] explored the challenges of detecting disinformation generated by LLMs due to the increased quality and effectiveness. Feng et al. [6] analyzed the potential of LLMs for bot detection and the associated risks, finding that LLM-based detectors reached state-of-the-art accuracy. However, they also showed that the existing detectors yield less accuracy on LLM-enhanced bot content.

Prior work pointed out that most bot detection research focuses on isolated tools, recommending future studies to prioritize standardized definitions, unsupervised methods, and applicability across different types of bots [4]. In addition, most prior research focused on Twitter as opposed to other forms of social media. We extend our methodology to other social media platforms as they have gained traction among our target demographic of college students. Additionally, user adaptation to the detection tools remains unknown. Our study explores patterns from users' adaptation to the current presence of bots on social media and the detection strategies adopted by them, which can inspire future detection techniques.

3 METHODS

3.1 Data Collection

We conduct our study through semi-structured interviews. Firstly, because the interview process gives us the opportunity to acquire in-depth responses capturing the various methods that users adopt to detect bots along with the rationale behind those methods. Follow-up questions allow us to capture details about participants' perception of bots and individual experiences that might not be captured through a survey composed of a fixed set of questions.

3.1.1 Recruitment. Considering the recent studies revealing that the social media usage is highly prevalent among college students [13] and the limited time and monetary requirements of this project, we targeted University of Maryland students as our participants. In order to eliminate any bias that can be caused by participants' expertise on bots or detection strategies, we restricted our participation criteria to students from non-computing majors. We defined non-computing majors as a major that is not listed as a computing major by the UMD Student Success Office [12]. While we understand that many students who are not in computing majors may have strong technical backgrounds that would improve their ability to detect bots on social media, it is less likely to find bot detection experts in this population given our limited ability to recruit participants.

We recruited participants by messaging students we personally know, by posting and advertisement on the UMD Reddit page, and University of Maryland student organizations that we are members of, and preparing fliers to distribute in person. Unfortunately, our Reddit post was removed by Reddit's content filters. To enforce our requirement for participants to not be in a computing major, we invited prospective participants to complete a pre-screening survey before being able to schedule an interview. The survey asked for the prospective participant's field of study as well as their age, the social media platforms they use (among Reddit, Facebook, X, and Instagram), the amount of time they spend on social media, and the purpose of their social media use. The full survey responses are displayed in Table 2.

3.1.2 Interview Procedure. After selecting the participants based on their responses in the pre-screening questionnaire, we interviewed five students in-person at the Brendan Iribe Center for Computer Science and Engineering compensating them $15 for their time. The interviews were conducted with one researcher being present each time, and the subsequent audio recordings and transcriptions were performed using Zoom's AI Companion software.

Table 1: Scenarios Discussed in Interviews

Scenario Type	Scenario
Passive Scenario	User observing reply to a post made by someone else
Passive Scenario	User observing a post in their feed made by another account
Active Scenario	User receiving a reply to their own post
Active Scenario	User receiving message from customer support bot

We split the interview into two main sections: questions about social bot awareness, questions about user detection strategies. The first section explores how users define social bots, their history of interacting with bots, and their opinions of bots. The second section explores the detection strategies adopted by users under four scenarios: two passive detection scenarios and two active detection scenarios. We define a passive scenario as one where a user passively observes posts that are independent of their account on social media: in our case users reading replies from other users to a post that has not been posted by themselves, and a post on their feed. Active scenarios are ones where the users are allowed to engage in an active communication with the potential bots: when users read replies to their posts or receive direct messages and communicate with a customer support account. We also investigate users' openness towards detection tools with the questions assessing their prior experience with detection tools, and their openness towards and expectations from such tools. The specific scenarios employed during the interviewing process are displayed in Table 1. A detailed outline of the exact questions used can be found in Appendix B.

3.2 Data Analysis

To qualitatively code our interview results, we derived a list of keywords and specific strategies identified in the interviews. To do so, we used a shared Google Sheet to construct a codebook of recurring themes for each question in the interview. Each researcher coded the results of their own interview, and a second researcher reviewed each coding for accuracy. We ensured consistency across coding by using the same Google Sheet, allowing the creation of a shared codebook.

For the first section of the interview (social bot awareness) we simply looked for the prevalence of recurring themes across the five participants. For the second section (user strategies) we compiled a list of every strategy the users suggested across the passive and active detection questions. We used this list to report trends across the users and make suggestions for bot detection algorithms and tools.

3.3 Limitations

The small sample size and focus on UMD students severely limits our study's external validity. Anecdotally, college students use social media very frequently with interactions with classmates, friends, and student groups being done online. Therefore, they may be more likely to have exposure to and experience with social bots than older users. Additionally, since our sample is small and recruitment was primarily over word-of-mouth to personal contacts, the diversity of our population is not reflective of the diversity of the general population. All of the participants reported using Instagram as their primary social media platform, and they have limited experience with the other platforms. Future research in user bot detection strategies should address these shortcomings by recruiting a more representative sample.

3.4 Ethical Considerations

This study was conducted with careful attention to ethical guidelines related to participant privacy, consent, and data security. All participants were voluntarily recruited from the University of Maryland, provided with a $15 incentive for their participation, and gave written informed consent before their interviews began. Participants were fully informed of the purpose of the study, the nature of the questions, and their right to withdraw at any time without negative effects. To ensure confidentiality, the interview recordings were deleted once transcription was completed and verified, and all personal identifiers were removed from the transcripts. The study was reviewed and approved by the Institutional Review Board (IRB) at the University of Maryland, ensuring that all procedures adhered to ethical standards for research involving human subjects.

Beyond participant welfare, ethical considerations also extend to the broader implications of this research. As the study explores strategies for detecting social bots, the findings could influence the development of future bot-detection tools. While these tools aim to support user agency and improve platform transparency, they may also raise privacy and surveillance concerns if not designed responsibly. To address this, researchers advocate for the development of detection systems that prioritize user control, ease of use, transparency, and explainability. These principles ensure that users remain informed about how detection decisions are made, reducing the risk of misuse or harm.

4 RESULTS

4.1 Social Bot Awareness

Our participants agreed that social bots are non-human accounts. P1 gave a broad definition, saying *"they're basically just isn't a person behind [a bot] account."* P2 added that the accounts are *"maintained by robots...but they are mimicking human behavior."* P4 gave a narrower definition thinking specifically of features like *"Meta AI, or on Snapchat the AI feature."* When asked about their feelings of social bots, three participants explicitly stating feeling uncomfortable with social bots. Of the two remaining participants, one (P3) shared they were concerned about younger and older users recognizing the bots, and that they were "annoying", while the other (P2) reported that they *"don't know"* but *"wouldn't actually want to follow a social bot."*

The participants generally reported infrequent interactions with social bots. P2 reported the lowest frequency of "never." P1 and P3 mentioned they interacted with social bots "once every 3 to 4 months" and "very rarely" respectively. P4 and P5 reported the greatest frequencies of "2 to 3 times per month" and "about once a

Item	P1	P2	P3	P4	P5
Field of Study	Marketing	Second language Acquisition, PhD	Psychology	Finance & Philosophy	Aeronautical Engineering
Age	19	32	21	19	20
Most used Social Media platform	Instagram	Instagram	Instagram	Instagram	Instagram
Average daily time spent on Social Media	1 to 2 hours	30 minutes to 1 hour	less than 30 minutes	less than 30 minutes	2 to 3 hours
Main purpose of Social Media	Other (all options apply)	Professional networking	Online shopping	Connecting with friends and family	Entertainment

Table 2: Participant Responses to the Pre-Screening Survey.

month" respectively. The most common reported bot interactions were bots tagging the participant in fake giveaways (P1, P3, P4), a bot commenting on a post asking to draw or paint the participant (P1, P3, P4), and bot accounts offering payment over direct messages for their time (P1, P3, P4). P4 and P5 reported times where their friends' accounts were compromised and were used to scam them. P2, having never interacted with a bot online, did not report any experiences.

When asked how they observe online communities responding to perceived bot posts, the majority of participants mentioned they observed users calling out bot posts (P1, P3, P4, P5) and that they treat bots as jokes (P1, P3, P4, P5). The fifth participant (P2) said they observe users following bot accounts for fun.

Interestingly, the participants unanimously agreed that it is "very important" to be able to detect bots. When asked how comfortable participants would feel unknowingly interacting with bots, four participants (P1, P2, P3, P5) reported that they would be very bothered , while P4 *"wouldn't really mind terribly"* if they found out they were interacting with a bot. When asked about the positive and negative impacts of bots on social media, three of the participants (P1, P3, P5) explicitly said bots make no positive impact, and P2 said that the negative impacts outweigh the positive impacts. P4 said they *"don't really see a positive side of it, besides being able to make jokes about it."* The participants generally expressed negative sentiment and discomfort towards online social bots, and shared nothing positive other than being able to share a laugh with other users about them.

4.2 User Bot Detection Strategies

4.2.1 Passive Detection Strategies. For the first scenario, asking how the user would detect a bot if they replied to a post on social media, the participants listed several strategies they would employ to detect bots. Four participants (P2, P3, P4, P5) stated they would look for language that does not resemble human writing in posts. Three participants explicitly mentioned they would look at the account's profile to make a determination of whether it is a bot (P1, P3, P4). Participants shared that profile elements like really basic usernames (P1, P3) and blurry profile photos (P1, P5) in particular are likely associated with bots. They also shared that in addition to non-human language, posts that include advertisements (P1, P5), requests to draw a mural of a user (P1), incorrect scientific concent (P4) and content that lacks emotional langua P2) are associated with bot accounts.

For the second scenario, asking how users would detect a bot if they saw a bot's post on their feed, the users shared some additional strategies they would look for. P1 shared several additional strategies for detecting bots, such as checking if the posting account's username reads as a typosquatted version of their friend's username, determining if they posted an unrealistic number of photos over a short time, and having a relatively young account. P4 added that they would check the profile's bio to determine if it is "unrealistic." P2 mentioned that an unreasonably high number of followers is another possible feature of a bot. P5 added that grainy photos of generic landmarks suggests copy/pasted content, which is associated with bots. Across both scenarios, the participants generally agreed that observing an account's behavior in addition to just the language used in a specific post is useful for determining if the account is a bot. We list the reported per-participant passive detection strategies in Table 3 of Appendix B.

4.2.2 Active Detection Strategies. For the first scenario, asking how users would detect a bot if they saw a potential bot reply to their post, the participants reported additional strategies to actively probe the potential bot to determine if they are a bot. Three participants explicitly stated that they would try to provoke user actions by replying to the potential bot and analyzing its reaction (P1, P3, P5). P2 mentioned they would ask personal questions like "How do you feel?" to determine if they are human. One participant, P4, did not suggest any strategies involving active interaction, but added they would check to see if the potential bot wrote the same message to other users.

For the second scenario, asking how users would respond to a customer support account, the participants similarly reported they would ask questions to determine if the account is human. Four participants reported that bots would give unclear answers with asked specific questions (P2, P3, P4, P5). Two participants said they would ask very complicated questions and see if the account replies unrealistically quickly (P2, P3). One participant misinterpreted the question as asking if they received a message from a bot impersonating a customer support account rather than a deliberate interaction

with a customer service account (P1). P1 reported that in this case, the account asking for passwords or personal information while the user is already logged in is very suspicious. Across both scenarios, participants agreed that asking specific and sentimental questions to analyze how an account responds is useful for determining if the account is a bot. We list all the reported passive detection strategies and which participants reported them in Table 4.

4.2.3 Impact of Demographics. An interesting avenue to pursue would be to explore if a user's field of study and background with social media influences their strategies for detecting bots. We observed that P1 and P2 listed the most strategies for passive scenario and active scenario bot detection respectively. P1 is a marketing major and said during the interview they *"get paid to do work for social media right now, like I manage a company social media"*, which may suggest a relationship between a user's professional use of social media, their field of study and the strategies they derive. Given the limited number of participants and the diversity in their academic backgrounds, any attempt to generalize connections between field of study and bot detection strategies would be speculative. Our results suggest that a larger, more diverse participant pool would be required to assess this relationship in a statistically meaningful way.

4.2.4 Confidence in Detection. For each scenario, we asked the participants how confident they would be on detecting bots under given scenario. Results present an increasing confidence under active communication scenarios in comparison to the passive scenarios due to ability of crafting messages to reveal information about the bot and the occurrence of new evidence such as the timing of the replies during an active conversation. Specifically, for detection of costumer support bots, all participants expressed a high or moderate confidence. P2 compared customer support bots to ChatGPT: *"Best social bot that I'm in interacting every day is ChatGPT, it's really good. It's not like talking to a customer support service"*, suggesting that the users are aware of the capabilities of advanced generative models, but they do not associate the underlying technology of advanced chat bots to customer support bots.

4.2.5 Bot Detection Features and Tools. When asked about bot detection tools, all five participants reported that they are unaware of any existing bot detection tools, but would be open to their use. Three participants stated they do not feel bots are enough of a problem to justify using a tool (P1, P2, P3). They agreed that they would only use a tool if it is convenient (P1, P3, P4, P5) and low-cost (P1, P2, P5). This general lack of interest in the tool is ironic considering how all the participants expressed concerns and discomfort about social bots in the previous section.

When asked what features bots could implement to resist detection, participants suggested that using a normal-looking profile photo (P1, P3), a biography that makes sense (P1, P3, P5), emotional language (P2) and more formal language (P5) makes bots harder to detect. Additionally, if bots respond to users more adaptively (P3, P4) and vary their posting patterns so it looks human (P1), they become harder to detect. Notably these suggestions are direct countermeasures to several of the strategies users mentioned.

5 DISCUSSION AND CONCLUSION

This study provides a comprehensive analysis of how non-expert users detect social bots on social media platforms. Our findings reveal substantial variability in user definitions of social bots, with participants describing them as "AI users," "fake profiles," and "scammers." This definitional ambiguity underscores the need for more user education on what constitutes a bot, as misaligned definitions may hinder effective detection. While users' detection strategies differ between passive and active scenarios, their confidence was notably higher in active cases, where direct engagement with bots allowed for interactive testing of responses. This insight highlights the potential for future bot-detection tools to incorporate active detection mechanisms, empowering users to probe and verify suspected accounts.

Participants' reliance on visual and behavioral cues for bot detection highlights opportunities for system design. For passive detection, users emphasized features like blurry profile photos, basic usernames, and generic posts as key signals. Active detection, on the other hand, involved users asking targeted questions to assess the contextual awareness of responses. These human-driven strategies suggest that detection tools could benefit from surfacing similar cues automatically, such as flagging suspicious usernames or encouraging user-initiated "challenge questions" during interactions. Importantly, our study found that users underestimated the frequency of bot interactions, with most participants believing they rarely encounter bots putting the total number of perceived social media bot interactions at 1-3 per month, and one participant stating that number is closer to zero. Various recent studies suggest these numbers are likely much higher in practice, with a 2017 study estimating the percentage of bot accounts on X between 9% and 15% [18], and a more recent 2020 work putting this estimate firmly at 15% [14]. This misperception may stem from the growing sophistication of AI-driven bots, underscoring the need for tools that bring unseen interactions to users' attention.

While participants acknowledged the potential utility of bot-detection tools, demand for such tools was low. Users were only willing to adopt a tool if it was free, convenient, and required minimal user effort. This finding suggests that bot-detection features would be most effective if integrated directly into social media platforms as "background" services. For designers, this implies a shift toward passive, low-effort interactions where visual indicators of bot-like behavior are subtly displayed. Additionally, the results suggest that hybrid detection models—leveraging user-driven signals like username quality and system-driven signals like content analysis—could bridge the gap between manual and automated detection.

Our study offers timely insights for the design of human-centered bot-detection systems. By highlighting the differences in user strategies across passive and active scenarios, we outline a path for more interactive, user-driven detection tools. Nevertheless, limitations such as the small sample size (n=5) and recruitment from a single university suggest caution in generalizing the findings. Future work should explore larger and more diverse participant pools, examine platform-specific strategies, and investigate hybrid detection models that blend human intuition with AI-driven classification. As social bots continue to grow in sophistication, the development of

more accessible, context-aware tools will be essential for fostering safer and more transparent social media environments.

REFERENCES

[1] Dennis Assenmacher, Lena Clever, Lena Frischlich, Thorsten Quandt, Heike Trautmann, and Christian Grimme. 2020. Demystifying Social Bots: On the Intelligence of Automated Social Media Actors. *Social Media + Society* 6, 3 (2020), 2056305120939264. https://doi.org/10.1177/2056305120939264 arXiv:https://doi.org/10.1177/2056305120939264

[2] Chiyu Cai, Linjing Li, and Daniel Zengi. 2017. Behavior enhanced deep bot detection in social media. In *2017 IEEE International Conference on Intelligence and Security Informatics (ISI)*. 128–130. https://doi.org/10.1109/ISI.2017.8004887

[3] Zi Chu, Steven Gianvecchio, Haining Wang, and Sushil Jajodia. 2010. Who is tweeting on Twitter: human, bot, or cyborg?. In *Proceedings of the 26th Annual Computer Security Applications Conference* (Austin, Texas, USA) *(ACSAC '10)*. Association for Computing Machinery, New York, NY, USA, 21–30. https://doi.org/10.1145/1920261.1920265

[4] Stefano Cresci. 2020. A decade of social bot detection. *Commun. ACM* 63, 10 (Sept. 2020), 72–83. https://doi.org/10.1145/3409116

[5] Clayton Allen Davis, Onur Varol, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. 2016. BotOrNot: A System to Evaluate Social Bots. In *Proceedings of the 25th International Conference Companion on World Wide Web* (Montréal, Québec, Canada) *(WWW '16 Companion)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 273–274. https://doi.org/10.1145/2872518.2889302

[6] Shangbin Feng, Herun Wan, Ningnan Wang, Zhaoxuan Tan, Minnan Luo, and Yulia Tsvetkov. 2024. What Does the Bot Say? Opportunities and Risks of Large Language Models in Social Media Bot Detection. arXiv:2402.00371 [cs.CL] https://arxiv.org/abs/2402.00371

[7] Robert Gorwa and Douglas Guilbeault. 2018. Unpacking the Social Media Bot: A Typology to Guide Research and Policy: Unpacking the Social Media Bot. *Policy Internet* 12 (08 2018). https://doi.org/10.1002/poi3.184

[8] Antonio Pequeno IV. 2024. *Musk's X Says It's Purging Bots—Here's How The Platform Has Struggled To Squash Its Bot Problem.* https://www.forbes.com/sites/antoniopequenoiv/2024/04/04/musks-x-says-its-purging-bots-heres-how-the-platform-has-struggled-to-squash-its-bot-problem/ Forbes, April 4, 2024.

[9] Bohan Jiang, Zhen Tan, Ayushi Nirmal, and Huan Liu. 2023. Disinformation Detection: An Evolving Challenge in the Age of LLMs. arXiv:2309.15847 [cs.CL] https://arxiv.org/abs/2309.15847

[10] Sneha Kudugunta and Emilio Ferrara. 2018. Deep neural networks for bot detection. *Information Sciences* 467 (2018), 312–322. https://doi.org/10.1016/j.ins.2018.08.019

[11] Mehwish Nasim, Andrew Nguyen, Nick Lothian, Robert Cope, and Lewis Mitchell. 2018. Real-time Detection of Content Polluters in Partially Observable Twitter Networks. In *Companion Proceedings of the The Web Conference 2018* (Lyon, France) *(WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1331–1339. https://doi.org/10.1145/3184558.3191574

[12] UMD Student Success Office. [n. d.]. Computing Majors. https://studentsuccess.umd.edu/computing-majors

[13] OpenR. 2022. The Growing Use of Social Media Among College Students. https://openr.co/the-growing-use-of-social-media-among-college-students/#google_vignette Accessed: 2024-12-09.

[14] Jorge Rodríguez-Ruiz, Javier Israel Mata-Sánchez, Raúl Monroy, Octavio Loyola-González, and Armando López-Cuevas. 2020. A one-class classification approach for bot detection on Twitter. *Computers Security* 91 (2020), 101715. https://doi.org/10.1016/j.cose.2020.101715

[15] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. 2018. The spread of low-credibility content by social bots. *Nature Communications* 9, 1 (Nov. 2018). https://doi.org/10.1038/s41467-018-06930-7

[16] Stefan Stieglitz, Florian Brachten, Björn Ross, and Anna-Katharina Jung. 2017. Do Social Bots Dream of Electric Sheep? A Categorisation of Social Media Bot Accounts. arXiv:1710.04044 [cs.HC] https://arxiv.org/abs/1710.04044

[17] Ruixiang Tang, Yu-Neng Chuang, and Xia Hu. 2023. The Science of Detecting LLM-Generated Texts. arXiv:2303.07205 [cs.CL] https://arxiv.org/abs/2303.07205

[18] Onur Varol, Emilio Ferrara, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2017. Online Human-Bot Interactions: Detection, Estimation, and Characterization. *Proceedings of the International AAAI Conference on Web and Social Media* 11, 1 (May 2017), 280–289. https://doi.org/10.1609/icwsm.v11i1.14871

[19] Yuhao Wu, Yuzhou Fang, Shuaikang Shang, Jing Jin, Lai Wei, and Haizhou Wang. 2020. A Novel Framework for Detecting Social Bots with Deep Neural Networks and Active Learning. https://doi.org/10.1016/j.knosys.2020.106525

[20] Kaicheng Yang and Filippo Menczer. 2024. Anatomy of an AI-powered malicious social botnet. *Journal of Quantitative Description: Digital Media* 4 (May 2024). https://doi.org/10.51685/jqd.2024.icwsm.7

[21] Kai-Cheng Yang, Emilio Ferrara, and Filippo Menczer. 2022. Botometer 101: social bot practicum for computational social scientists. *Journal of Computational Social Science* 5, 2 (Aug. 2022), 1511–1528. https://doi.org/10.1007/s42001-022-00177-5

[22] Sarita Yardi, Daniel Romero, Grant Schoenebeck, and danah boyd. 2010. Detecting Spam in a Twitter Network. *First Monday* 15 (01 2010). https://doi.org/10.5210/fm.v15i1.2793

APPENDIX A - PRE-SCREENING SURVEY

The pre-screening survey asked the following questions:

- What is your field of study? Please respond with your major(s) and minor(s)
- How old are you?
- Which of the following social media platforms do you use? (Select all that apply)
  - Reddit
  - Facebook
  - X
  - Instagram
- What is the average time you spend daily on social media?
  - Less than 30 mins
  - 30 mins to 1 hour
  - 1 to 2 hours
  - 2 to 3 hours
  - More than 3 hours
- What is your purpose when using social media?
  - Connecting with friends and family
  - Following news and current events
  - Entertainment (videos, memes, etc.)
  - Professional networking
  - Sharing content (posts, photos, videos, etc.)
  - Online shopping
  - Participating in online communities or discussions
  - Learning new skills or information
  - Other (write-in)

APPENDIX B - INTERVIEW QUESTIONS

The interview asked the following questions split over two sections.

5.1 Social Bot Awareness Questions

(1) Do you know what a social bot is? If so, how would you describe it?
(2) How do you feel about the presence of human-like bots on social media?
(3) Have you ever noticed that you are interacting with a bot? (reading the posts of a bot, or a bot directly replied to you)
  - If yes, what was the context of the interactions? (reading the posts of a bot etc)
  - If yes, what were these bots trying to do? (promoting a political opinion, advertising etc, entertainment)
  - If yes, how frequent does it happen?
  - Are there any scenarios that you can think of, in which a bot can interact with you online?
(4) Have you ever seen others discussing bots in online communities (e.g., in comments, threads)? How do people generally react to bots in those spaces?

(5) What types of social media platforms do you think are more prone to bot activity, and why?
(6) How much would it bother you to interact with a bot without realizing?
(7) Do you think it's important to be able to confidently detect bots on social media? Why/why not?
(8) What positive and negative impacts can bots provide on social media?

5.2 User Strategy Questions

(9) (Passive Scenario) If you were reading replies to a post on the platform that you use the most (amongst X ,Reddit, Instagram and Facebook), how would you tell if any of those replies were created by a bot or a real person?
- If this happened on one of the other platforms (X, Reddit, Instagram, Facebook), would your approach to identifying the bot be different? How so?
- How would the detection strategy change based on the content (political, personal, cultural, entertainment, advertisement)
- How confident would you be about your strategy?

(10) (Passive Scenario) If you were scrolling through the feed of the platform you use the most (amongst X ,Reddit, Instagram and Facebook), what features would raise your suspicion that a post's owner is a bot?
- If this happened on one of the other platforms (X, Reddit, Instagram, Facebook), would your approach to identifying the bot be different? How so?
- How would the detection strategy change based on the content (political, personal, cultural, entertainment, advertisement)
- How confident would you be about your strategy?

(11) (Active Scenario) If you received a reply from an unknown source to a post on social media, how would you tell if it is a bot or real person?
- How confident would you be about your strategy?
- Suppose you continued the conversation, how would you structure your messages to help determine if the account is managed by a bot?

(12) (Active Scenario) If you received a response from a customer support account after contacting a company, how would you determine whether it is a human or a bot managing the account?
- How confident would you be about your strategy?
- Suppose you continued the conversation, how would you structure your messages to help determine if the account is managed by a bot?

(13) In your opinion, what features of a bot would make it more difficult to identify?
(14) Are you aware of or have you ever heard of any tools that could help you detect bots online?
- Have you ever used such a tool?
- How open would you be to using such tools in the future?

Table 3: Passive Detection Strategies Mentioned by Participants

Strategy	P1	P2	P3	P4	P5
Participant would look at an account profile	X		X	X	
A blurry photo of a girl's face as the profile photo makes bots detectable	X				X
Bots have really basic usernames	X		X		
Would be weary of bot advertisements	X				X
Bots ask to draw photos/murals of users	X				
Bots often use non-human language		X	X	X	X
Participant would fact-check the content itself if scientific/research				X	
Participant would check for lack of emotinal language		X			
If the content is more knowledge-based, the facts may created by a bot		X			
Participant would check for lack of personal opinion		X			
Participant would look for grammar mistakes (makes it less likely to be bot)		X			
An account using a friend's name with an extra dash/character is likely a bot	X				
Participant looks for posting patterns	X				
An account posting an unrealistic number of photos over a short time is likely a bot	X				
Account age is a factor (newer accounts are more likely to be bots)	X				
Bots have silly or unrealistic bios in their profile				X	
Having a high number of followers makes an account more likely to be a bot		X			
Posting grainy, low quality pictures of famous things/memes is associated with bots					X

Table 4: Active Detection Strategies Mentioned by Participants

Strategy	P1	P2	P3	P4	P5
Participant looks at content of replies to determine if it was posted by a bot	X		X		
Participant are mainly concerned with asking an account questions to get a reply, and then determine if the account is a bot	X		X		X
Participant would ask personal questions		X			
Participant would send meaningless messages and inspect the reaction	X				X
Participant would check if a reply was sent to multiple posts				X	
Participant would look for reoccurring expressions		X	X		
Participant Would look for a neutral, formal tone		X			
Participant feels that if the sentence is too perfect, the account can be a bot"		X			
Replies that look to come from someone who doesn't know the participant are likely bot-generated					X
It is suspicious for a legimiate account to ask for passwords on the platform when user is logged in	X				
An account using the chat feature of social media to provide technical support is suspicious	X				
Participant would ask questions to verify legitimacy of tech support	X				
Bots respond with FAQ/non-answers		X		X	
Tech support leaving unclear answers to specific questions is indicative of bots		X	X	X	X
Participant would ask a question that is time-consuming to answer		X	X		
Profile photo not matching the name is indicative of bots	X				
Language that is too formal is indicative of bots		X			
Repeating expressions is indicative of bots		X			

Table 5: Participant Themes in Question 1: Do you know what a social bot is? If so, how would you describe it?

Theme	P1	P2	P3	P4	P5
Bots use fake profile information	X				X
Bots are computer-controlled	X	X			
Bots include AI social media users	X		X		
Bots scam users	X			X	X
Bots include built-in AI features like snapchat AI and Meta AI			X	X	

Table 6: Participant Themes in Question 2: How do you feel about the presence of human-like bots on social media?

Theme	P1	P2	P3	P4	P5
Participant feels uncomfortable with bots on social media	X			X	X
Younger users have trouble recognizing bots	X		X		
Older users have trouble recognizing bots	X		X		
Bots try to scam victims	X				X
Bots engage in false advertising	X				X
Participant finds bots not interesting		X			

Table 7: Participant Themes in Question 3: Have you ever noticed that you are interacting with a bot? (reading the posts of a bot, or a bot directly replied to you)

Theme	P1	P2	P3	P4	P5
Bots have tagged the participant in a fake giveaway	X		X	X	
Participant observed a bot tagging others in fake giveaway	X		X		
Bots have commented about drawing/painting a user	X		X	X	
Bots sometimes pretend to be a duplicate account of a friend	X				
Bots sometimes are "Sugar Daddy" accounts and offer payment for time	X		X	X	
Males get fake DMs from bots masquerading as beautiful girls	X		X		
Bots compromised a friend's account and used it for attacks				X	X

Table 8: Participant Themes in Question 4: Have you ever seen others discussing bots in online communities (e.g., in comments, threads)? How do people generally react to bots in those spaces?

Theme	P1	P2	P3	P4	P5
Users mess with the bot pubicly	X		X		
Users call out bots	X		X	X	X
Participant believes bot comments stand out a lot	X		X		X
People treat bots as jokes	X		X	X	X
There is a "negative connotation" with the bots			X	X	
People interact with (follow) bots for fun		X			

Table 9: Participant Themes in Question 5: What types of social media platforms do you think are more prone to bot activity, and why?

Theme	P1	P2	P3	P4	P5
Sites with lots of people are prone to bots	X		X		X
Sites with older victims are prone to bots	X				
Sites with less tech savvy participants are prone to bots	X				
There are more bots on Instagram than on Reddit	X		X		
Sites with teenage victims are prone to bots	X				
Sites with more advertising are more prone to bots	X	X	X		X
Direct messaging features are more of a concern w.r.t. bots than other social media			X	X	
Sites that emphasize entertainment are more prone to bots		X	X		

Table 10: Participant Themes in Question 6: How much would it bother you to interact with a bot without realizing?

Theme	P1	P2	P3	P4	P5
Participant would be very bothered by the bot interaction	X	X	X		X
Participant would be mad at themself later	X		X		X
Participant would wonder how the bot's profile is realistic enough for them to fall for it	X				
Participant would not interact willingly with bots unless it's a known bot site	X		X		
Participant would feel their ego busted	X				
Participant would not "mind terribly"				X	
Participant would feel deceived		X	X		

Table 11: Participant Themes in Question 7: Do you think it's important to be able to confidently detect bots on social media? Why/why not?

Theme	P1	P2	P3	P4	P5
It is very important to be able to detect bots	X	X	X	X	X
Participants want to avoid divulging sentitive info	X	X	X		
Participants do not want to fall for scams	X		X	X	X
Participant fings bot detection important for privacy reasons		X			

Table 12: Participant Themes in Question 8: What positive and negative impacts can bots provide on social media?

Theme	P1	P2	P3	P4	P5
Participant explicitly said there are no positives w.r.t social bots	X		X		X
Participant explicitly said there are no negatives w.r.t social bots					
The negatives of bots can outweigh positives		X			
Bots can scam people (negative)	X			X	X
Bots take money and trick you (negative)	X		X		
Can make jokes (positive)				X	
Bots defeat the purpose of social media (negative)					X
Passive Detection					

Table 13: Participant Themes in Question 9: If you were reading replies to a post on the platform that you use the most (amongst X ,Reddit, Instagram and Facebook), how would you tell if any of those replies were created by a bot or a real person?

Theme	P1	P2	P3	P4	P5
Participant would look at an account profile (strategy)	X		X	X	
Parcipant believes bots are set up explicitly for deception	X				
A blurry photo of a girl's face as the profile photo makes bots detectable (strategy)	X				X
Bots have really basic usernames (strategy)	X		X		
Participant would feel less confident on X	X			X	
Participant would feel less confident on Reddit	X			X	X
Some platforms allow links in comments for bots to use	X				
Would be weary of bot advertisements (strategy)	X				X
Bots ask to draw photos/murals of users (strategy)	X				
Participant feels they can just tell a bot reply from human	X		X		
Bots often use non-human language (strategy)		X	X	X	X
Participant would fact-check the content itself if scientific/research (strategy)				X	
Participant not be that confident overall with detection	X			X	X
Participant would check for lack of emotional language (strategy)		X			
If the content is more knowledge-based, the facts may created by a bot (strategy)		X			
Participant would check for lack of personal opinion (strategy)		X			
Participant would look for grammar mistakes (makes it less likely to be bot) (strategy)		X			
Participant mentioned how good ChatGPT is on mimicking human-like language		X			

Table 14: Participant Themes in Question 10: If you were scrolling through the feed of the platform you use the most (amongst X ,Reddit, Instagram and Facebook), what features would raise your suspicion that a post's owner is a bot?

Theme	P1	P2	P3	P4	P5
Bots have really basic usernames (strategy)	X		X		
A blurry photo of a girl's face as the profile photo makes bots detectable (strategy)	X				
An account using a friend's name with an extra dash/character is likely a bot (strategy)	X				
Participant believes giveaway accounts are purchased	X				
Participant looks for posting patterns (strategy)	X				
An account posting an unrealistic number of photos over a short time is likely a bot (strategy)	X				
Account age is a factor (newer accounts are more likely to be bots) (strategy)	X				
Bots have silly or unrealistic bios in their profile (strategy)				X	
Having a high number of followers makes an account more likely to be a bot (strategy)		X			
Posting grainy, low quality pictures of famous things/memes is associated with bots (strategy)					X
Active Detection					

Table 15: Participant Themes in Question 11: If you received a reply from an unknown source to a post on social media, how would you tell if it is a bot or real person?

Theme	P1	P2	P3	P4	P5
Participant looks at content of replies to determine if it was posted by a bot (strategy)	X		X		
Participant are mainly concerned with asking an account questions to get a reply, and then determine if the account is a bot (strategy)	X		X		X
Participant would just mess with the bot if they know it is a bot	X				
Participant would ask personal questions (strategy)		X			
Participant would send meaningless messages and inspect the reaction (strategy)	X				X
Participant would check if a reply was sent to multiple posts (strategy)				X	
Participant would look for reoccurring expressions (strategy)		X	X		
Participant Would look for a neutral, formal tone (strategy)		X			
Participant feels that if the sentence is too perfect, the account can be a bot (strategy)"		X			
Replies that look to come from someone who doesn't know the participant are likely bot-generated (strategy)					X

Table 16: Participant Themes in Question 12: If you received a response from a customer support account after contacting a company, how would you determine whether it is a human or a bot managing the account?

Theme	P1	P2	P3	P4	P5
It is suspicious for a legimiate account to ask for passwords on the platform when user is logged in (strategy)	X				
An account using the chat feature of social media to provide technical support is suspicious (strategy)	X				
Participant would ask questions to verify legitimacy of tech support (strategy)	X				
Bots respond with FAQ/non-answers (strategy)		X		X	
Tech support leaving unclear answers to specific questions is indicative of bots (strategy)		X	X	X	X
Participant would ask a question that is time-consuming to answer (strategy)		X	X		

Table 17: Participant Themes in Question 13: In your opinion, what features of a bot would make it more difficult to identify?

Theme	P1	P2	P3	P4	P5
Normal profile photos would make bots harder to detect	X		X		
Profile bios that look normal would make bots harder to detect	X		X		X
Bots posting over reasonable timeframes would make them harder to detect	X				
Profile photo not matching the name is indicative of bots (strategy)	X				
More adaptive responses would make bots harder to detect			X	X	
Emotional language would make bots harder to detect		X			
Language that is too formal is indicative of bots (strategy)		X			
Repeating expressions is indicative of bots (strategy)		X			
Formal language makes bots harder to detect					X

Table 18: Participant Themes in Question 14: Are you aware of or have you ever heard of any tools that could help you detect bots online?

Theme	P1	P2	P3	P4	P5
Participant is not aware of bot detection tools	X	X	X	X	X
Participant would be open to use if needed	X	X	X	X	X
Participant does not need the tool since bots arent enough of a problem	X	X	X		
Participant would use detector if using social media for work and money involved	X				X
Participant believes convenience is important	X		X	X	X
Participant doesn't want to have to type in the username to check an account	X				
Participant would like a red flag/visual graphic	X				
Participant would use tool only if cheap/free	X	X			X
Participant has general distrust in systems				X	
The tool should not hinder the performance of the app				X	
The tool should not ask for personal information		X			