You are a Bot! - Studying the Development of Bot Accusations on Twitter

Twitter, Social Bots, Accusations, Dataset, Social Media Analytics

Extended Abstract

The use of social media platforms has increased dramatically over the last decade, with Twitter being one of the most popular platforms worldwide. However, the presence of bots on social media has become a growing concern due to their potential to manipulate public opinion and influence important events, such as elections. Although there have been several studies on the detection and characterization of bots, little attention has been paid to the impact of bots, and the accompanying phenomena on platform users and society. In this data-driven study, we address this research gap by analyzing a novel dataset of bot accusations on Twitter since 2007. We focus on the evolution of bot accusations over time and explore users' perception of the construct "bot" at a large scale.

Within our work, we answer the following research questions:

- How did bot accusations on Twitter change over time?
- What are the contexts in which Twitter users accuse others of being a bot?
- Do the definitions of bots internalized by Twitter users align with the definitions used in popular bot detection methods?

For data acquisition, we focus on Reply-Tweets that contain explicit bot accusations. We collected as many candidate accusation situations as possible in the first phase of data collection, accepting a higher number of false positives to achieve high recall. In the second phase, we filtered out instances deemed irrelevant to our research design, aiming to reduce false positives while maintaining high precision. We used the Twitter v2 API's full-archive search endpoint and a query of *"bot is:reply lang:en"* to match all English Tweets containing the keyword "bot" that were sent as a reply to another Tweet. After data retrieval, we were left with 35,876,388 Tweets that matched our query and were considered potential accusation situations. We filtered these down by only retaining the situations where the accusing Tweet contained the regular expression *"you are a [a-z]*bot|you're a [a-z]*bot"* to increase precision.

Our analysis reveals that the term "bot" has evolved from its technical meaning to become an "insult" used in polarizing discussions to discredit and dehumanize the opponent. Furthermore, our study sheds light on the contexts in which bot accusations occur and the motivations behind them. We also compare the definitions of bots internalized by Twitter users with those used in popular bot detection methods and find some inherent discrepancies that evolved over time.

Overall, our study provides insights into the impact of bots on social media platforms from the user's perspective. By exploring the evolution of bot accusations over time and the contexts in which they occur, we contribute to a better understanding of how users perceive bots and their potential to affect public discourse. We hope that our study will motivate future research to examine the effects of bot accusations on individual users and society, and to redirect research efforts towards studying the impact of bot accusations on public discourse rather than only focussing on the development of bot detection methods.

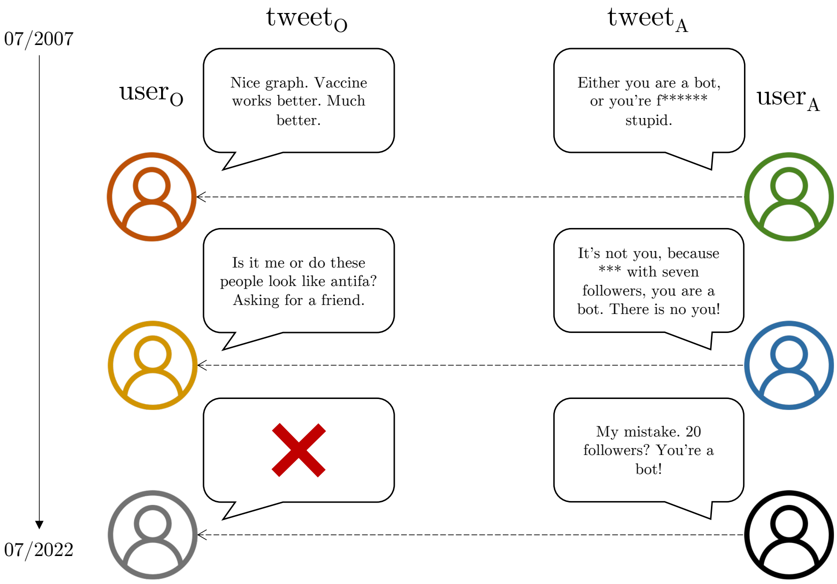
tweet$_O$
tweet$_A$
07/2007
user$_O$
Nice graph. Vaccine works better. Much better.
Either you are a bot, or you're f****** stupid.
user$_A$
Is it me or do these people look like antifa? Asking for a friend.
It's not you, because *** with seven followers, you are a bot. There is no you!
My mistake. 20 followers? You're a bot!
07/2022

Figure 1: Overall structure of our new dataset. We always capture accusation pairs. The dataset is structured to include various accusation situations over time. While the objects $user_A$ (accusing user) and $tweet_A$ (accusing Tweet) are consistently present in all cases due to the data collection strategy, the objects $user_O$ (accused/original user) and $tweet_O$ (accused Tweet) may be missing as they are not retrievable by the API anymore

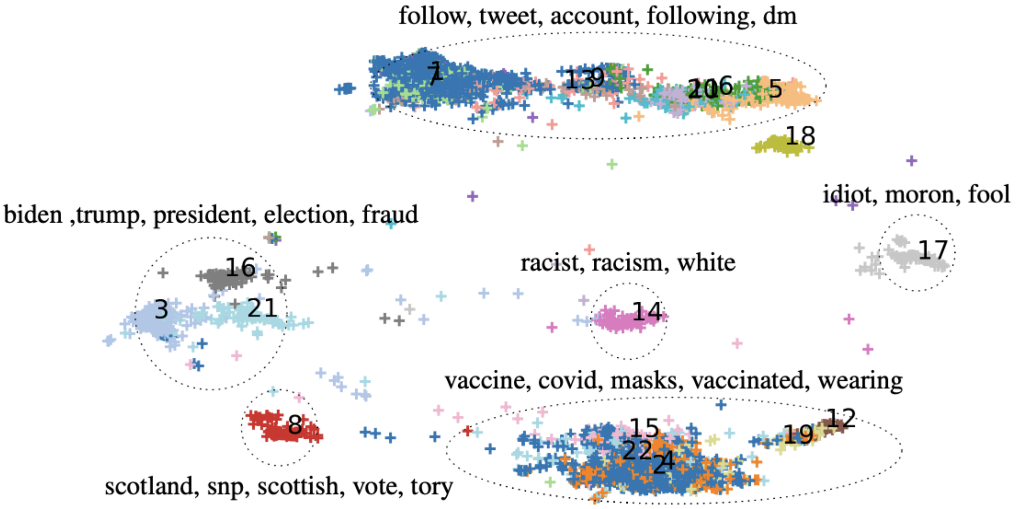

Figure 2: The 2021 Tweets of $user_O$, who is accused of being a bot, have been projected onto a 2-dimensional space, and top clusters have been identified and labeled with the highest class-based-TFIDF terms. In contrast to previous years, the accusations against users are no longer limited to automated behavior, but are specifically related to polarizing debates such as covid/mask/vaccine, election/Biden/Trump, Scottish independence, or racism. Furthermore, a cluster containing only insults has been observed, indicating an increase in toxicity surrounding bot accusations.