scientific reports

OPEN

A global comparison of social media bot and human characteristics

Lynnette Hui Xian Ng & Kathleen M. Carley

Chatter on social media about global events comes from 20% bots and 80% humans. The chatter by bots and humans is consistently different: bots tend to use linguistic cues that can be easily automated (e.g., increased hashtags, and positive terms) while humans use cues that require dialogue understanding (e.g. replying to post threads). Bots use words in categories that match the identities they choose to present, while humans may send messages that are not obviously related to the identities they present. Bots and humans differ in their communication structure: sampled bots have a star interaction structure, while sampled humans have a hierarchical structure. These conclusions are based on a large-scale analysis of social media tweets across ~ 200 million users across 7 events. Social media bots took the world by storm when social-cybersecurity researchers realized that social media users not only consisted of humans, but also of artificial agents called bots. These bots wreck havoc online by spreading disinformation and manipulating narratives. However, most research on bots are based on special-purposed definitions, mostly predicated on the event studied. In this article, we first begin by asking, "What is a bot?", and we study the underlying principles of how bots are different from humans. We develop a first-principle definition of a social media bot. This definition refines existing academic and industry definitions: "A Social Media Bot is An automated account that carries out a series of mechanics on social media platforms, for content creation, distribution and collection, and/or for relationship formation and dissolutions." With this definition as a premise, we systematically compare the characteristics between bots and humans across global events, and reflect on how the software-programmed bot is an Artificial Intelligent algorithm, and its potential for evolution as technology advances. Based on our results, we provide recommendations for the use of bots and for the regulation of bots. Finally, we discuss three open challenges and future directions of the study of bots: Detect, to systematically identify these automated and potentially evolving bots; Differentiate, to evaluate the goodness of the bot in terms of their content postings and relationship interactions; Disrupt, to moderate the impact of malicious bots, while not unsettling human conversations.

Keywords Social media bot, Psycholinguistic analysis, Social identity, Network analysis

The notion of "bots" on social media is ubiquitous across many scholarship. These studies captured a range of different social phenomena where bots operate: politics, hate speech, toxicity and so forth. Bots were used to boost the follower count of politicians in the 2011 Arab Springs uprising, generating false impressions of popularity[1,2]. In the same uprising, bots flooded news streams to interrupt efforts of political dissidents[1,2]. In the US 2020 elections, bots augmented human users in strategic communications, and actively distorted or fabricated narratives to create a polarized society[3–5]. More recently, bots aggressively pushed anti-vaccine narratives and conspiracy theories on social media during the 2021 coronavirus pandemic[6,7]. Bots applied social pressure to influence humans to favor the anti-vaccine ideology[3,8]. When the tension of the online ideologies are sufficiently strong, and the spread sufficiently wide, these ideologies spillover to the offline world, resulting in protests, riots and targeted hate-speech[9–12]. Social media bots gained further media attention in 2022 when Elon Musk proclaimed that at least 20% of the Twitter users were bots, which were influencing content quality[13]. Musk later bought the platform, and took steps to curtail the bot population in a "global bot purge", which includes removing huge amounts of bots, and charging newly created accounts to post and interact on the platform[14].

Much research on social media bots involve constructing bot detection algorithms and applying bot detection algorithms to analyze bot activity during an event. Bot detection algorithms typically extract a series of features from user and post data, then build a supervised machine learning model which classifies the likelihood of a

Center for Computational Analysis of Social and Organizational Systems, Societal and Software Systems Carnegie Mellon University, Pittsburgh, PA 15213, USA. email: lynnetteng@cmu.edu

user being a bot or a human[15]. These machine learning models range from logistic regression[16], to random forests[17], to ensemble of classifiers[4,18], to deep learning methods[19,20]. Another technique of bot detection is graph-based methods, which infers the probability of a user being a bot by its connections, i.e. friends[21,22]. Most recently, Large Language Models (LLMs) are incorporated in bot detection algorithms to handle the diverse user information and content modalities[23]. These bot detection classifiers have been used to study bot behavior in many events, ranging from political events[4,24–26] to natural disasters[27,27] to the spread of information and opinions on social media[8,28,29].

Although researchers have built automatic bot detection classifiers, behavioral studies show that humans are unable to differentiate between the bot and human user[30]. In fact, the identification of bots by security students are akin to random guesses[31]. Consequently, it is important to study bots and their characteristic nature and activity patterns. Our study is positioned within the social cybersecurity realm, which studies how the digital environment, particularly bots, can be exploited to alter the content and community relationships[32].

This article is being driven by looking at a first principles approach to bots. We ask the following research questions:

- **RQ1: What is a social media bot?** Many studies are predicated on a general purpose understanding of a bot, or a specific definition particular to the event of study. Instead, we pare down the definition of a bot into its treatment of the core elements of a social media platform (users, content, relationships).
- **RQ2: How does the nature of a bot differ from a human?** Systematically, we look at the difference between bots and humans. We use a large scale dataset from Twitter/X that spans over 200 million users, and analyzed four aspects: volume of bot/human user types, use of linguistic cues, use of identity terms, and social interactions.

After an examination of the social media bot, we discuss how a bot is also an Artificial Intelligent (AI) agent, and its potential evolution alongside technological advancements. We finally follow with a discussion of the open research challenges of the study of bots to encourage future studies in this field. The challenges we identify reflect the definition of a bot: Detect, to systematically identify these automated and evolving bots; Differentiate, to evaluate the goodness of the bot in terms of their content postings and relationship interactions; and Disrupt, to moderate the impact of malicious bots, while not unsettling digital human communities.

What is a social media bot?

The term "bot" has become a pervasive metaphor for inauthentic online users[8,33]. Most social media users have an implicit understanding of a bot, as do most researchers[30]. Table 1 summarizes some of the recent definitions of bots retrieved from academic literature. The security industry also watches social media bot accounts, and Table 2 summarizes definitions from industry sources. Each of the definition grasps one or more relevant properties (highlighted in bold) of a social media bot, yet are not sufficiently comprehensive to describe the bot. Some of these relevant properties are: "automated", "interacts with humans", "artificial agents".

Year	References	Definition
2016	33	A social bot is a computer algorithm that **automatically produces content** and **interacts with humans** on social media, trying to emulate and possibly alter their behavior.
2016	26	[...] social bots, **algorithmically controlled accounts** that **emulate the activity of human users** but operate at much higher pace (e.g., automatically producing content or engaging in social interactions), while successfully keeping their artificial identity undisclosed
2016	50	**Automated accounts**, called bots, [...]
2018	58	Bots are have been generally defined as **automated agents** that function on an online platform [..]. As some put it, these are programs that run continuously, formulate decisions, act upon those decisions without human intervention, and are able adapt to the context they operate in.
2018	69	The term "social bot" describes accounts on social media sites that are **controlled by bots** and **imitate human users** to a high degree but differ regarding their intent.
2018	17	[...] malicious **automated** agents
2020	20	Social Media Bots (SMB) are **computer algorithms** that **produce content** and **interacts with users**
2020	38	[...] social bots, **(semi-) automatized** accounts in social media, gained global attention in the context of **public opinion manipulation**.
2020	18	Malicious actors create **inauthentic social media accounts** controlled in part by **algorithms**, known as social bots, to disseminate misinformation and agitate online discussion.
2021	70	Social bots - partially or fully **automated accounts** on social media platforms [...]
2022	71	Social media bots are **automated accounts** controlled by software algorithms rather than human users
2023	41	Social bots are **automated** social media accounts **governed by software** and controlled by humans at the backend.
2023	15	A bot is a software that mimics human behavior and **operates autonomously and automatically**.
2023	72	Twitter accounts controlled by **automated programs**.
2023	73	**Automated accounts** on social media that **impersonate real users**, often called "social bots,"
2023	74	Social bots are social media accounts **controlled by software** that can **carry out content** and **post content** automatically.
2024	30	Social bots are **artificial agents** that infiltrate social media
2024	75	Social bots are social media accounts **controlled in part by software** [...] Social media bots display profiles and **engage** with others through various means, including following, liking, and retweeting

Table 1. Definitions of "Social Media Bot" in academic literature.

Year	References	Definition
2018	US Department of Homeland Security[39]	[...] Social Media Bots as programs that vary in size depending on their function, capability, and design; and can be used on social media platforms to **do various useful and malicious tasks while simulating human behavior**
2024	Microsoft[76]	Social media bots are **automated programs** designed to **interact** with account users.
2024	Meltwater[77]	Refers to the definition by US CSIA (see below)
Not Dated	CloudFlare[37]	[...] social media bots are **automated programs** used to **engage** in social media. These bots behave in an either partially or fully **autonomous fashion**, and are often designed to **mimic human users**.
Not Dated	Cybersecurity and Infrastructure Security Agency (CISA)[44]	Social Media Bots are **automated programs** that **simulate human engagement** on social media platforms.
Note Dated	Imperva[78]	An Internet bot is a software application that runs **automated tasks** over the internet.

Table 2. Definitions of "Social Media Bot" in industry literature.

One of the problems with existing definitions is that they often define bots as being malicious and they highlight the nefarious use of bots[5,34–37]: "malicious actors", "public opinion manipulation", "malicious tasks"[18,38,39]. Most often, the study of bots is established upon nefarious tasks: election manipulation, information operations, even promoting extremism[33,34,40]. The exact same technology can be used in both good and bad ways. There are plenty of good bots[41–43]. Bots provide notifications and entertainment[44], such as the @CoronaUpdateBot found in our dataset which posts critical public health information. Bots support crisis management efforts by gathering the needs and combined locations of people after a disaster, for authorities and community volunteers to identify crucial areas and providing help[45]. Chat bots provide emotional support during stress[46] and continue bonds in times of grieve[47]. Celebrities and organizations use bots to facilitate strategic communications with their fans and clients[3,48].

In essence, a bot is a computer algorithm. As an algorithm, a bot is neither bad or good. It is the persona , or public facade, that the bot is afforded to that determines the goodness of its use. We develop a generic definition of a bot. The definition is independent of the use of the bot. The determination of the use of the bot warrants separate treatment beyond this paper. Regardless of whether a bot is used for good or ill, the behavioral characteristics of bots remain the same.

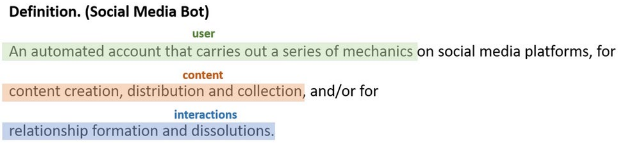
To better describe the social media bot, we first need to characterize the environment in which it lives: the social media platform. Within a social media platform, there are three main elements: users, content and relationships[49]. Users are represented by their virtual accounts, and are the key actors driving the system, creating and distributing information. Content is the information generated by users on the platform. Relationships are formed from the user-user, user-content and content-content interactions.

After distilling a social media platform into its core building blocks, it follows that definition of a social media bot should be based on the foundations of a social media platform as first principles. The presence of these components in each of the reference definitions are broken down in Table 3. The first principles of a bot are:

- User: *An automated account that carries out a series of mechanics.* A key characteristic of bots is its programmability, which give it its artificial and inauthentic characteristic. Automation is an aspect that has been iterated in all the reference definitions. The key here is that a bot is automated. The model underlying the automation does not matter; any model can be applied equally well to humans and bot. A bot could be built to mimic humans, or it could be built to optimize other functions. For example, some studies describe bots in terms of its mimicry of humans[33,50], but others observe that bots eventually influence the social space such that humans mimic bots[51,52]. The automated nature of the bot gives rise to its systematic nature, which bot detection algorithms leverage through machine learning algorithms. In contrast, human users create accounts for a wide number of reasons, from personal to business to pet accounts, and have very varied and random behavior mechanics.
- Content: *for content creation, distribution, and collection and processing.* Bots often generate their content in bulk to distribute a certain ideology[24,53], such as a good portrayal of their affiliated country[54]. Instances where bots perform content distribution is where the spread fake news and disinformation content[28,55,56], or when they distribute general news information[57]. Bots in the form of web crawlers and scrapers download and index data from social media in bulk[58], and sometimes process the data to perform functions like analyzing sentiment of opinions[59]. Humans create and distribute content to create their online personalities, to update their friend groups of their lives, or to promote their businesses[60,61]. However, humans seldom generate or collect content in bulk; if they do so, they are likely to use a bot to aid them, rendering their account a cyborg[3].
- Relationships: *for relationship formation and dissolution.* In other words, forming a relationship online means to connect with other users via post-based (i.e., retweet, mention) or user-based (i.e., following, friend) mechanisms. Dissolving a relationship means to destroy a connection by forcing users to leave a community. Bots are an actively form and destroy relationships on social media platforms. An example of the formation of post-based relationship is the active amplification of narratives. This technique is mostly employed in the political realm where the bots retweet political ideology in an organized fashion[24,62,63]. User-based relationships can grow through coordinated fake-follower bots, that are used to boost online popularity[64], or can be dissolved through toxic bots that spread hate and directed religious ideologies[40,65], causing users to break away from the community[66].In general, bots form and dissolve automated relationships towards some single-purpose goal, while humans form and dissolve relationships organically, sometimes based on real-life encounters.

	User		Content		Interactions	
References	Automation	Mimicry	Creation	Distribution	Communication	Relationship
[33]	x	x	x			
[26]	x	x	x		x	x
[50]	x					
[58]	x					
[69]	x	x		x		
[17]	x					
[20]	x		x		x	x
[38]	x					
[18]	x			x		
[70]	x					
[71]	x					
[15]	x	x				
[41]	x	x	x			
[72]	x					
[73]	x	x				
[74]	x		x	x		
[30]	x					
[75]	x				x	x
US Department of Homeland Security	x	x				
Microsoft	x				x	x
CloudFlare	x	x			x	x
CISA	x	x			x	x
Imperva	x					

Table 3. Components of definitions of "Social media Bot".


Definition. (Social Media Bot)

user
An automated account that carries out a series of mechanisms on social media platforms, for
content
content creation, distribution and collection, and/or for
interactions
relationship formation and dissolutions.

Fig. 1. Definition of Social Media Bot. This definition displays the possibilities of mechanics that the bot account can carry out. A bot does not necessarily carry out all the mechanics.

Type of bot	Use for good	Use for bad
General Bot	Search engine optimization, data collection[58]	Spread disinformation[29], manipulate opinion[8]
Bridging Bot	Political commentators that aggregate information[16]	Disseminate information to instigate anger across social, cultural, community differences[54]
Political Bot	"Establishing brand and amplifying messages"[3,25], Digital campaigning[16]	Political manipulation[67]
Chat Bot	Emotional support during stress[46] and grieve[47]	Post offensive and inflammatory comments[68]
Activist Bot	Crisis management[45]	Trigger and initiate activism[2,9]

Table 4. Illustration of type of bots and their role in the social media space. Note that this list is not exhaustive but illustrative.

Figure 1 reflects a first principles definition of a social media bot. A Social Media Bot is "An automated account that carries out a series of mechanics on social media platforms, for content creation, distribution, and collection and processing, and/or for relationship formation and dissolutions." This definition displays the possibilities of mechanics that the bot account can carry out. A bot does not necessarily carry out all the mechanics. The combinations of mechanics that a bot carries out thus affects the type of bot it is and the role it plays within the social media space, and as shown in Table 4, those mechanics can be used for either good or bad. Bot types can be named for their actions or for their content. For example, a bot that carries out relationship formation between two different communities, and does not do any content mechanics can be classified as a bridging bot[16].

We illustrate a few types of bots and their use for good and bad in Table 1. Note that this list is not meant to be an exhaustive list but an illustrative list of the variety of bots in the social media space.

Results

We perform a global comparison of bot and human characteristics by combining several datasets obtained from X (previously named Twitter) using the Twitter V1 Developed API. These events are: Asian Elections[25,34], Black Panther[79], Canadian Elections 2019[80], Captain Marvel[81], Coronavirus,[82] ReOpen America[9,82] and US Elections 2020[82]. In total, these datasets contain ~ 5 billion tweets and ~ 200 million users. Each user in this database is labeled as bot or human using the BotHunter algorithm[17].

We used a suite of multidisciplinary methods for our analysis. This is illustrated in Fig. 2. We integrated machine learning classification, social network analysis and linguistic feature extraction in a robust hybrid methodology that bridged computational, linguistic and network sciences. We analyzed the behavior of bots and humans along four axes. We analyzed the behavior of social media users along four axes. The first is classifying users into bot or human using a machine learning model called BotHunter. The second was linguistic feature analysis of tweets and user metadata. These information were extracted from the NetMapper software, then averaged across user type (i.e., bot vs human) and event type, and compared using statistical tests. The third is the self-presentation of social identities in the user metadata. These identities are derived from matching with an occupation census. In this third analysis, we also extracted topic frames that were written in the tweet texts using the NetMapper software. Thereafter, we correlated the social identities with the topic frames, linking together user presentation with user writing. Fourth, we construct interaction networks of users using the ORA software, and analyzed the influence of bots and humans in terms of their position and structure of their ego networks. Further details about our methodology and implementation can be found in the Supplementary Material.

How many bots are there?

Figure 3 presents the percentage of bot users within each dataset. On average, the bot volume across the events are about 20% with the bot percentage spiking up to 43% during the US Elections. This is in line with past work, where a general sample of users usually reveal a bot percentage below 30%[72], yet in a politically-charged topic (i.e. elections, tensions between countries), the bot percentage rises[34,82]. Our estimate is also empirically consistent with Elon Musk's estimate of 20%[13]. This finding is important for event analysis, because it provides a comparison baseline towards the percentage of bot-like users within an event. Spikes in bot user percentage beyond 20% suggest that the event and conversation has caught the interest of bot operators, and the analyst can monitor for signs of conversation manipulation.

How do bots differ from humans linguistically?

We extract psycholinguistic cues from the tweets using the NetMapper software[83]. The software returns the count of each cue in the sentence, i.e., the number of words belonging to the cue in the tweet. There are three categories of cues: semantic, emotion and metadata. Semantic and emotion cues are derived from the tweet text, while metadata cues are derived from the metadata of the user. Semantic cues include: first person pronouns, second person pronouns, third person pronouns and reading difficulty. Emotion cues include: abusive terms, expletives, negative sentiment, positive sentiment. Metadata cues include: the use of mentions, media, URLs,

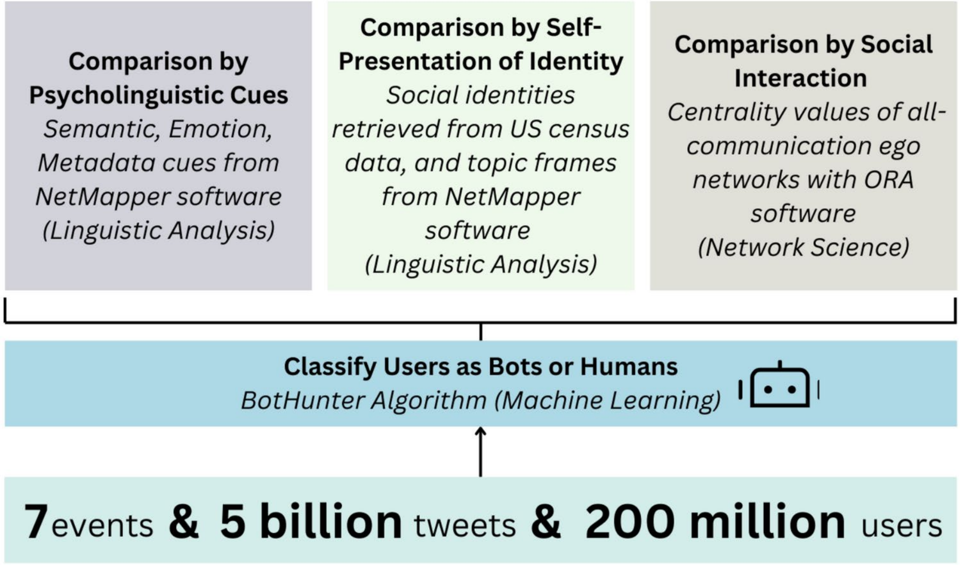

Comparison by Psycholinguistic Cues
Semantic, Emotion, Metadata cues from NetMapper software
(Linguistic Analysis)
Comparison by Self-Presentation of Identity
Social identities retrieved from US census data, and topic frames from NetMapper software
(Linguistic Analysis)
Comparison by Social Interaction
Centrality values of all-communication ego networks with ORA software
(Network Science)
Classify Users as Bots or Humans
BotHunter Algorithm (Machine Learning)
7 events & 5 billion tweets & 200 million users

Fig. 2. Illustration of multidisciplinary methods used for our analysis. We used the v1.0.82 for NetMapper software and v.3.0.9.181 for ORA software.

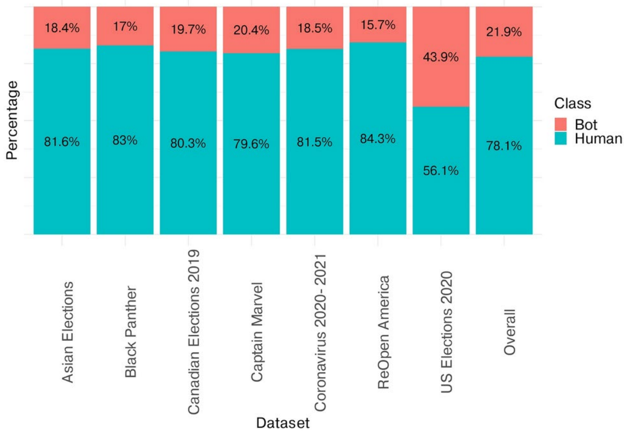


Fig. 3. Comparison of Bot volume across events. The percentage of bot users across the events are on average around 20%.

hashtags, retweets, favorites, replies, quotes, and the number of followers, friends, tweets, tweets per hour, time between tweets and friends:followers ratio.

Figure 4a presents the differences between cues used by bots and humans. The detailed numeric differences are in the Supplementary Material Table S2). This difference is examined overall, and by event. There are consistent differences in the use of cues by bots and humans. For example, across all events, bots use significantly more abusive terms and expletives, and tweet more than humans. On the other hand, humans use more first person pronouns, positive sentiment, and media (i.e., images, videos). Humans tend to quote and reply to tweets, while bots tend to retweet.

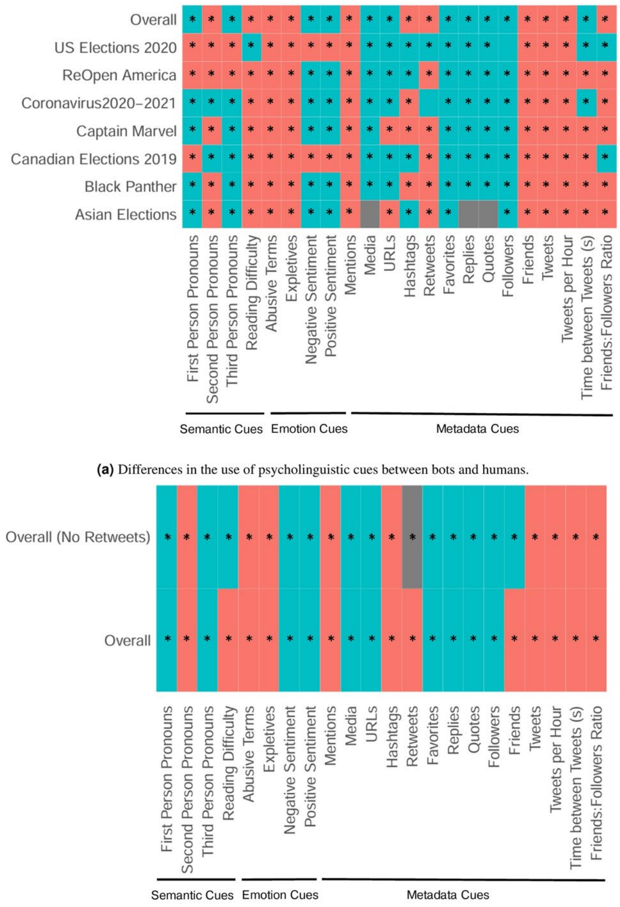
Most events have consistent cue distribution, but some events look different. In general, humans use more sentiment cues. However, in the two elections (US Elections 2020 and Canadian Elections 2019), bots used more sentiment cues. This reveals a deliberate attempt to use bots during the election seasons to polarize online sentiments. Prior research has shown that bots can be highly negative during the election season[84], and that bots express hugely different sentiment sentiment when mentioning different political candidates[8,85].

When a bot solely retweets a human, its linguistic profile, by definition, is identical to the human's. The question though, is whether the bots that are sending original tweets match the linguistic profile of those retweeting, or is the linguistic profile different? For the Black Panther and Captain Marvel events (Fig. 4b), we compared the psycholinguistic profile for all tweets, and the original tweets only (i.e., no retweets). In these two events, bots retweet significantly more than humans. In general, the bot-human difference between linguistic cue use of the original tweets vs all tweets are rather similar. However, the average tweet reading difficulty and the number of friends are different: in original tweets, humans have higher values; in all tweets, bots have higher values. Therefore, bots have their unique signature when generating new content (i.e., in original tweets), but are guaranteed to match human's content when retweeting the human's.

Bots construct tweets with cues that can be easily and heavily automated, while humans construct more personal tweets that require higher cognitive processing to create.This shows in Fig. 4a, where bots use more semantic cues, while humans use more emotion cues. For metadata cues, bots have more tweets and tweets per hour, while humans engage more with the conversation through replies and quotes. Such differences show how bot accounts still use rather rudimentary techniques: hashtag latching using multiple hashtags[27,40], connecting several users together with increased number of mentions[54] and flooding the zone with lots of tweets tweets of their desired narratives[24,86]. More sophisticated communication techniques like having an increased number of media, and more advanced interaction techniques that involve dialogue understanding like increasing the number of replies and quotes, are still left to the humans. In short, bots have not entirely mimicked humans, yet.

How do bots present themselves differently from humans?

Social identity theory depicts how social media users portray their image online, and the community that they want to be associated with[61,87]. We analyze the difference in the self-presentation of the identities between bots and humans, and the difference between the linguistic cues used by the identities. The results are presented in Fig. 5. Across the events, there are consistently a smaller proportion of bots that present with an identity. Overall, 21.4% of bots present an identity, while 27.0% of humans present an identity (see Supplementary Material Table S3). Bots are more likely to obfuscate their identities[88], allowing them to take on different personas to suit their operation requirements[89]. Figure 5a presents the top 25 identities by frequency between bots and humans. Overall, bots affiliated with a more limited set of identities (n=869 vs n=908 unique identities). The distribution of the identities for each dataset is presented in the Supplementary Material Tables S3 and S4. There is a larger drop in the frequency of the use of identities in bot users than in human users. For example, between the 3rd and 4th most frequently used identities, bots show a 33.2% decrease in the use of the 4th most frequent identity compared to the 3rd most frequent identities in bots, compared to a 30.0% decrease for humans. This



(a) Differences in the use of psycholinguistic cues between bots and humans.

(b) Differences in the use of psycholinguistic cues between bots and humans for the combination of Captain Marvel and Black Panther datasets. This compares the cue distribution with and without retweets.

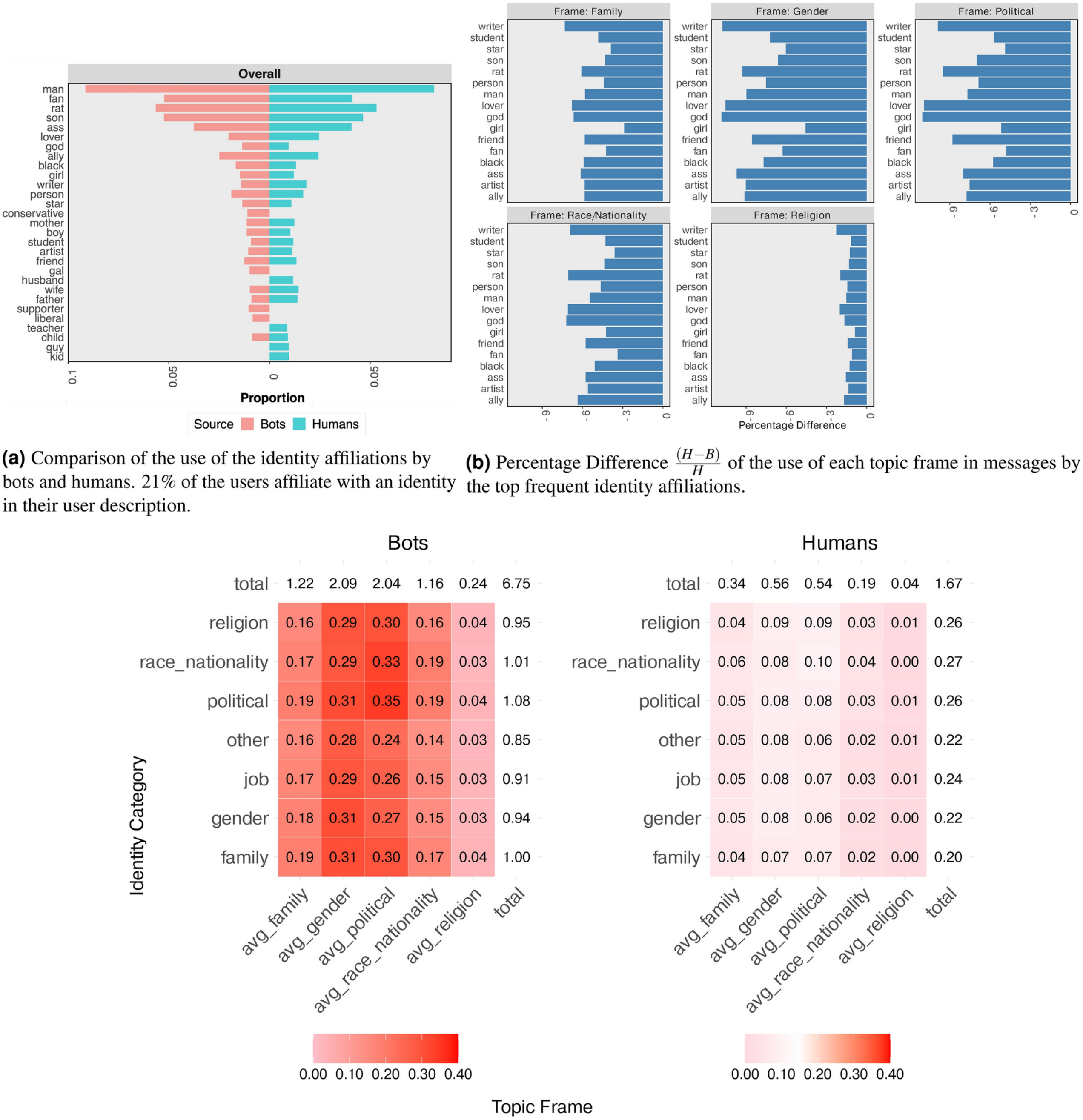
Fig. 4. Comparison of psycholinguistic overall cue usage (average cue usage per user) by bots and humans across datasets. **Green** cells show that **humans** use a larger number of the cue. **Red** cells show that **bots** use a larger number of the cue. * within the cells indicates there is a significant difference in the usage of the cue between bots and humans at the $p < 0.05$ level.

observation suggests that bots concentrate their self-presentation on certain identities, mostly the common ones: man, son, fan, lover; while humans have a more varied identity presentation.

We then ask a follow-up question: “How do the same bot/human identities talk about the same topics?” We compare the use of topic frames per identity for the most frequent identity affiliations in Fig. 5b. This frames are extracted from the NetMapper software (refer to Methods section for more details). Figure 5b plots the percentage difference of the use of framing cues (Family, Gender, Political, Race/Nationality, Religion) between bots and humans. This metric compares the use of cues with the human usage as a baseline. Overall, bots converse more aggressively in all topic frames. In particular, bots converse most around societal fault lines: gender, political, race/nationality. These conversations lie on societal fault lines, which could sow discord and chaos[90], therefore such bots are of interest to monitor and moderate. In fact, bots use more gender-based cues. Other research groups have also identified that a disproportionate number of bots that spread disinformation are females[91,92], and are thus more likely to use gender frames in their posts. Bots tend to converse largely about political topics, regardless of the identity they affiliate with, indicating that a good proportion of bots are deployed for political purposes, either by political parties or by political pundits[16,26,70]. Finally, the difference between the usage of topic frames between bots and humans could be due to their vocabulary used. The words used by humans are more varied and mostly not standard dictionary words, while bots are still being programmed with a limited set of vocabulary. This is evidenced by the proportion of words identified by the dictionaries in the NetMapper program used, where on average, humans use 4.98±2.45% more words related to topic frames compared to bots. The breakdown of percentage difference in words used is presented in the Supplementary Material Table S5. In a similar aspect, chat bot interactions have a more limited vocabulary than human interactions[93].

Figure 5c presents the average use of topic frames by identity categories. Humans affiliate themselves equally with all identity categories, while bots generally affiliate themselves with racial and political identities. Both bots and humans converse a lot on gender and political issues.

Bots converse mostly about topics that closely match their identity. For example, a bot that presents itself as “man” and “son” mostly converse about family then gender; while bots that take on the identities “conservative”



(a) Comparison of the use of the identity affiliations by bots and humans. 21% of the users affiliate with an identity in their user description.

(b) Percentage Difference $\frac{(H-B)}{H}$ of the use of each topic frame in messages by the top frequent identity affiliations.

(c) Average use of topic frames by identity category referred to. Bots are more likely to refer to gender and political identities, and are more likely to utilize racially typed language.

Fig. 5. Comparison of identity-related behaviors in bots and humans. The Methods section explains the derivation of the identity categories and topic frames. For detailed breakdown of identity-related behaviors per event, see Supplementary Material Figures S1, S2, S3, S4, S5, S6, S7, S8.

and "american" converse significantly more about politics. This observation can be read from the heatmap: for the bots that associate with the religion identity, the average use of religious words is 0.04, while that for humans is 0.00. If the users associate with the family identity, the average proportion of the use of family words within the content is 0.19 for bots and 0.04 for humans. Such is the curated presentation and programming of bot users, which allows for an aspect of predictability - if a bot user affiliates with a certain identity, it is likely to talk about topics related to its identity. This shows that bots are likely designed to look like humans. They are strategically designed to be in character by having the right affiliation to fit in and converse with a specific group.

Our observations in the affiliation of identities by bots in their user description and the use of identity-related topic frames means that bots are being used strategically. They are not just used to support or dismiss groups in general, but are specifically being aimed at a gender (i.e., women or men), or at a political actor (i.e., president, governor, politician). Bots are overused in the political, religious, and racial realm, suggesting that they are targeting topics of societal tensions.

How do bots communicate differently from humans?

Social interactions between users are an indication of the information dissemination patterns and the communication strategies of the users. We calculate the network metrics (total degree, in degree, out degree, density) of the all-communication ego-networks of the users. In the network graphs, the users are nodes, and the links between users represent all communications between the two users (i.e., replies, quotes, mentions, retweets). Table 5 compares the two metrics for bots and humans. Bot ego networks have higher density than human ego networks (8.33% more dense), which reflects that the bots have tighter communication structures and form more direct interactions than humans. On average, a bot has 9.66% bot alters and 90.34% human alters, whereas on average a human has 7.31% bot alters and 92.69% human alters. Although bots interact with a higher proportion of bot alters than humans do (32% more bot alters), our findings show that both bots and humans interact more with humans rather than bots in their ego network. By the principle of homophily, it is natural for humans to interact with other humans[94]. However, bots violate the principle of homophily, and instead of interacting with more bots, they interact with more humans. Therefore, bots are actively forming communication interactions with humans, perhaps attempting to influence humans[95].

Figure 6 shows the interaction of bots and humans in a network diagram. These users are illustrative of the most frequent communicators in the Asian Elections dataset. In this diagram, users are represented as nodes, and links between users represent a communication (e.g. a retweet, reply, mention). The network diagrams presented are one- and two-degree ego-networks, generated by the ORA software[83]. This means that the networks present users that are in direct communication with the user (1-degree), and are in direct communication with the 1st-degree users (2-degree).

A common way for bots to be used in political discourse (e.g. elections) is to amplify other users. As an amplifier, bots are the pendants of the user they are amplifying. Bots echo the ideas, narratives and emotions of these users, enhancing the visibility of the users and their narratives within the social network. Therefore, bots appear in star networks in many of the peripheral nodes. A star structure is a network that have a strongly connected core and peripheral networks. This structure is most prominent in bots in political discourse, where core bots create information, and peripheral bots amplify the information through the retweeting mechanic[24]. Humans, on the other hand, are more likely to be part of a tree structure, where one can make out the tiered first- and second-hop interactions. In the same discourse, humans are more likely to be performing many actions, sometimes retweeting other users, sometimes tagging other users and so forth.

This difference in interactions between bot and human users reveals the communication patterns of both user classes. The star structure of bots suggests that they have a hierarchy of interconnected bot users in an operation network to disseminate information, which is easily achieved with the help of automation. On the other hand, humans communicate predominantly within their immediate network before extending their communication outwards. The bot ego networks are more dense, signifying that they were constructed to interact more than do humans, and are sometimes constructed as networks of bots (the botnet)[95,96].

Discussion

Through our large scale empirical study, we show that bots and humans have interesting and consistent differences between them. These differences span from their volume, to the linguistic features of their text, to the identities they affiliate with, to their social network structure. These features can be used to characterize a social media bot, and how it differs from humans.

Our work introduces a valuable dataset containing billions of tweets over four years. This dataset is a valuable resource for studying patterns in social media bot technology. With the current restrictions by social media platforms for data collection, this dataset will be prohibitively expensive and technically not feasible to replicate. X API, this dataset will be extremely costly to collect. Using the free tier API, this dataset would take 50 million months or 136,986 years to complete collection. The Pro tier API costs allows retrieval of 1 million tweets per month at a cost of $5000, leading to the total cost of $25 million over 5000 months, or 416 years for the collection to be completed. Table S11 in the Supplementary Material details the calculations of the number of months and the cost required to replicate this dataset under the 2025 pricing structure.

We study a huge amount of data dated from 2018 to 2021. These data show consistent differences over the years, which means that while bot technology do evolve, it does not evolve drastically. Moreover, the consistent

	Bot	Humans
In-degree	0.05 ± 0.08	0.02 ± 0.02
Out-degree	8E-4 ± 1.4E−3	1.6E-3 ± 3.3E−3
Total degree	0.15 ± 0.09	0.16 ± 0.11
Density	0.35 ± 0.06	0.034 ± 0.06
% bot alters	9.66 ± 2.98	7.31 ± 3.10

Table 5. Comparison of network metrics. For the in-degree, out-degree, total degree and density, we present the ratio of mean(metric) for agent type : max(metric) across all agents in the event. For detailed breakdown of metrics per event, refer to Supplementary Tables 6, 7, 8, 9, 10.

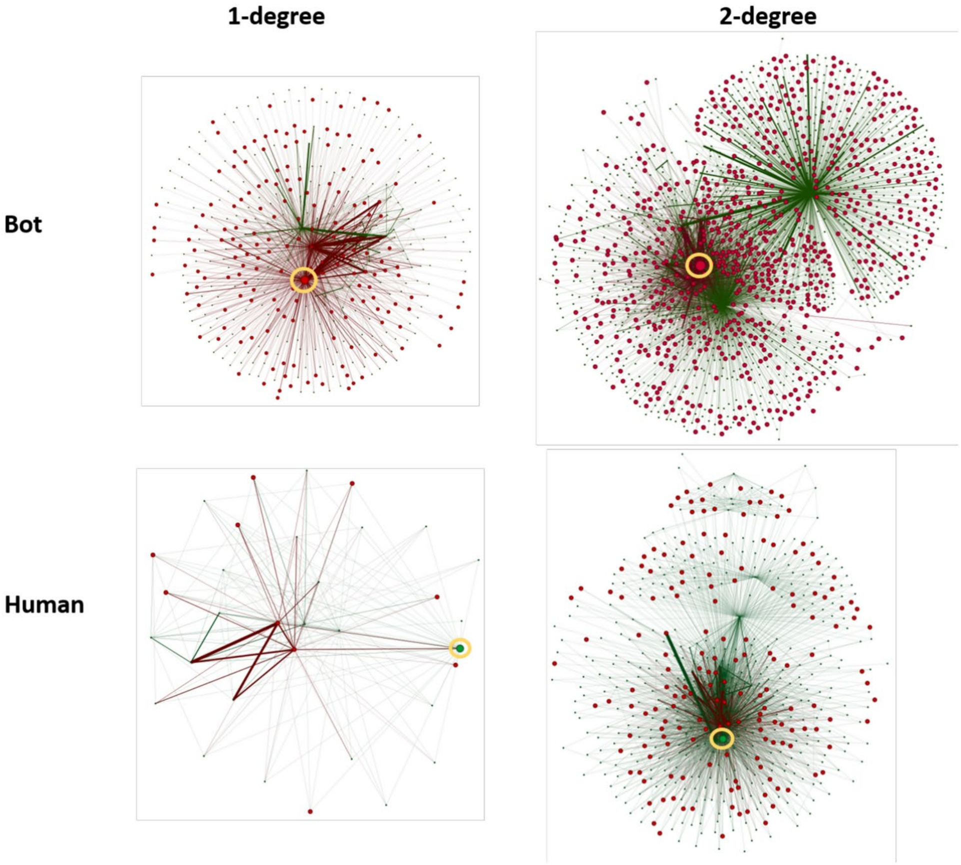


Fig. 6. Ego network structures of Bots and Humans who are the most frequent communicators in the Asian Elections dataset. Nodes represent social media users. Links between users represent a communication relationship between the two users (i.e., retweet, mention). Bot users are colored in red, human users in grey. The width of the links represent the extent of interactions between the two users. In these most frequent communicators, bots have a star network structure, and humans a tree structure. Bot networks have more bot alters, while human networks have more human alters.

behavioral differences show that there are scenarios where the automated nature of bots is more advantageous, and scenarios where the manual human control and thought is superior. These differences provide insights to how both can be utilized to afford conversations on social media: bots can be used for methodological postings with deliberate selection of hashtags, tweets per hour, and a structured star communication network. Humans can be used for more complex cognitive tasks such as adding media to a post or replying to a post, and for conversing on a larger range of topics[3].

Strengths of social media bots

An Artificially Intelligent (AI) system is a machine, or computer system, that can perceive its environment and use intelligence to perform actions to achieve defined goals[97]. The social media bot perceives the digital environment to decide their targets (i.e., users to retweet, users to mention), and intelligently carry out content and interaction mechanics to achieve their goals (i.e., spreading information[28,56], sowing discord[98,99], assisting the community[45–47]). The software-programmed social media bot is an AI algorithm, and thus has potential to be harnessed for social good.

Table 6 lists some recommendations of our results on how bots can be leveraged on for social good, based on possible avenues for interacting with the automated accounts. First, given that bots use more retweets and mentions than humans, and have high tweets per hour, bots can be used for menial tasks like announcements and distribution of information. Second, since bots have a star interaction network, they can be used for big announcements like disaster and crisis management without message distortion. A star network sends messages directly through interactions, hence the original message is preserved. However, the human's hierarchal interaction network will distort the message as it passes through the tiers. Third, bots typically post content that matches their identity, they can be used to provide educational material about topics that people associate with certain profession. For example, a weather news bot can provide weather information. Lastly, since bots use more abusive and expletive terms than humans, instead of regulating toxic language itself, regulation can be focused on disallowing bots to use such toxic language, which would therefore reduce the amount of hyperbole and

Result	Recommendation
Bots use a lot of retweets and mentions, and have high tweets per hour	Use bots for menial tasks like announcements and amplification of announcements
Bots have a star interaction network	Use bots for big announcements (e.g., disaster, crisis management) without message distortion
Bot content matches identity	Use bots to provide educational material about topics that people associate with certain professions (e.g. weather information from a weather news bot)
Bots use more abusive and expletive terms than humans	Focus regulation to disallow bots to use toxic language

Table 6. Recommendations of our observations on leveraging and regulating bots for social good.

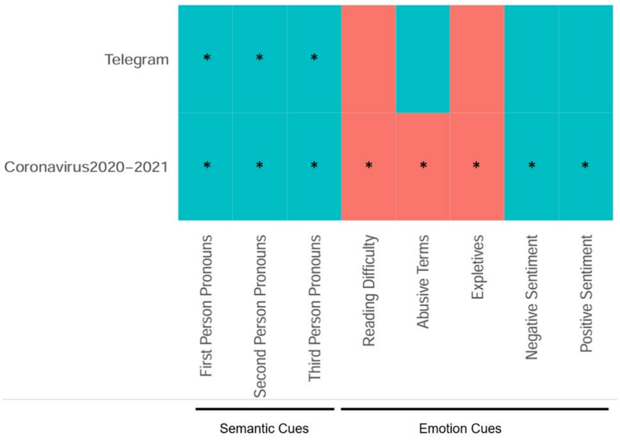


Fig. 7. Differences in the use of psycholinguistic cues between bots and humans for Twitter dataset and Telegram dataset. Both datasets are drawn from the Coronavirus2020-2021 event. **Green** cells show that **humans** use a larger number of the cue. **Red** cells show that **bots** use a larger number of the cue. * within the cells indicates there is a significant difference in the usage of the cue between bots and humans at the $p < 0.05$ level.

offense online. This regulatory recommendation draws on our observational study, and we suggest subsequent randomized controlled tests for validation before policy implementations.

Further investigations

In this paper, we primarily used an extensive dataset of Twitter data with tweets collected over four years. However, we have done other investigations to go beyond this dataset. In particular, we looked at bots across social media platforms and bot evolution.

Bots across Platforms Are bots the same across different social media platforms? To answer that question, we perform a preliminary analysis towards investigating the similarities and differences of bots on different social media platforms.

We compared the psycholinguistic behavior of bots between Twitter and Telegram platforms with respect to the Coronavirus 2020-2021 pandemic. The Telegram dataset contained $\sim$7 million messages written by $\sim$ 335,000 users, and was segregated into three user groups: bots, disinformation spreaders and humans[100]. We extracted psycholinguistic cues using the NetMapper software for the Telegram messages in the same fashion as we did for the Twitter messages (see Methods: Comparison by Psycholinguistic Cues for details).

Figure 7 presents the differences between bots and humans for a subset of the psycholinguistic cues. We only compare the semantic and emotion cues because the structure of the Telegram platform is different from the Twitter platform, so the metadata cues do not map directly. This comparison shows that the pattern of cue usage between bots are humans are similar across both platforms, yielding generalizability to our overall study. In the same light, past work had shown that there was insignificant difference in the user metadata and post structure between tweets from Twitter and messages from Telegram[100]. Bots are generally similar in the usage of linguistic cues across social media platforms, which provides wider generalizability of our results.

Bot Evolution Bots can evolve over time as adversarial actors realize that they are being detected and change their techniques[19]. Between 2021 and 2023, bots have enhanced their detection evasion techniques by adding four-character letter strings and Chinese proverbs to attempt to fool bot detection tools[101]. However, these techniques are not entirely novel. They are recombinations of social cybersecurity techniques like distract and distort techniques that were discovered in earlier studies[32,40]. These techniques revolve around permuting the actions that platforms provide for content and interactions.

Bots also continue to evolve as new technologies, such as Generative AI technologies, come out. We used three open-sourced Large Language Models (LLMs) to generate tweets related to the coronavirus event. These models are: Meta Llama 3.1 8B Instruct, Phi-4, and Qwen 2.5 7B Instruct 1M. Each model generated 20 tweets. The Supplementary Material details the methodology and the prompts used for this experiment in section *Generative-AI-based social media bots*, and shows the results of the BotHunter runs on the generated tweets in Supplementary Material Table S12 . Ideally, the use of generative AI technology would make the bot agents undetectable by current bot detection technology. That is, the use of LLMs would reduce the probability that the BotHunter algorithm classifies these tweets as originating from bot users. Across our experiments, the average BotHunter score retrieved was 0.69±0.15. This score is borderline on the 0.70 bot classification threshold that we used in this study, and is above the 0.50 threshold that many studies use[34,102–104]. The huge standard deviation indicates that LLMs generate personas that are inconsistent with a single user type: sometimes LLMs generated tweets that look like humans, sometimes they do not.

Despite the evolution of bots, bot detectors are still able to identify users as bots. The same BotHunter tool used in this study had been deployed on data from 2023 to 2025[16,105,106]. The tool classified 20% of the users as bot users, a percentage similar to that in our study. This shows that these bot detectors identify broad and comprehensive features. Moreover, these traditional bot types that do not employ generative AI still continue to exist, perhaps because the bot operator does not require generative technologies for his task. For example, a news aggregator bot that retweets a set series of news accounts can be programmed with heuristics. However, conversational bots will benefit from generative AI technologies and be able to carry out a more seamless conversation with the online community. Future work should examine the varied purposes of bots and whether evolution is required or whether evolution has occurred.

Future Work Future work involves more detailed investigation into the nuances of bot behavior across platforms. It is also worth exploring how the content layout and the relationship structure of different platforms affect the users' interaction dynamics. Future work can also take advantage of the temporal to analyze how bots evolve in behavior over time, and during crucial points in an event. The geographical dimension of our dataset further provides opportunity for a social geographical analysis on how bots differ across countries and regions.

Challenges and opportunities of studying social media bots

Next, we elaborate on three challenges in the study of social media bot, and discuss some opportunities for future research.

Detect

The first step to bot detection is to systematically detect these bots. However, these automated agents are constantly evolving and adapting their behavior in response to the changing setup of social media platforms and user patterns. The stricter data collection rules of social media platforms[107,108] and the increasing usage of AI in these bot agents[75] creates further variability in these digital spaces bots reside in. This therefore muddles any developed algorithms based on previous datasets.

Already, linguistic differences between bot and human tweets have narrowed between 2017 and 2020, making bot accounts more difficult to systematically differentiate[19]. More recently, AI-powered botnets have emerged, using ChatGPT models to generate human-like content[75], closing the gap between bot and human.

Bot evolution and bot detection are thus a "never-ending clash"[109], and sometimes bot accounts evolve faster than current known bot detection algorithms[70], presenting several opportunities in continual improvement of bot detection algorithms, specifically to be adaptable, faster, and more efficient. The increasing trends of using Large Language Models and Large Vision Models to create generated texts and deepfakes lend bots a helping hand in the construction of more believable narratives. These same generative technology are also used to construct offensive bots for humor[110]. However, current trends reflect that the use of such technologies is not very prevalent, for example,[75] only found one set of such botnet in their study, reflecting that bots are still relying on traditional techniques, likely because such heuristic-based techniques are easier and faster to deploy en masse.

Beyond simply detecting bot users in social media discourse, the next research agenda is to perform content-interaction impact analysis. Our work revealed one common interaction patterns bots use in their content communication. Hereupon arises an opportunity to characterize archetypes of interaction patterns, the best type of content to disseminate with each archetype, and the impact of the dissemination strategy on public opinion.

Differentiate

After identifying which users are likely to be bots, one must differentiate the goodness of the bot and its function. This evaluation can be inferred from the bot's content postings and relationship interactions. However, bots do not fall squarely in a spectrum of goodness; the lines of good and bad bots are blurred. In fact, bots can move between neutral in which they post messages that are not harmful, to bad, where they post politically charged and extremist messages[40,111]. Herein lies an opportunity to construct a rubric to determine the goodness of the bot; this, though, is a complex task, for there are ethical and societal issues to consider. Bots can change their goodness, too. They may be supporting a certain cause initially, then making a swing to a different stance soon enough. This swing of support was witnessed during the coronavirus pandemic era, where bots change their stances towards the vaccination campaign, disseminating narratives of different stances during different periods[8,85]. This opinion dynamics is important to the health of social discourse, and especially so when the bots require little conviction to change allegiances[8]. Another challenge involves identifying the type of bot, which can provide insight towards possible impact of the bot. For example, an Amplifier Bot that intensifies political messages could be intended to sow discord[24,63].

Disrupt

The third challenge is to mindfully disrupt the operations of bot users. That is, moderating the impact of malicious bots, while not unsettling human conversations. While banning bot users can be an easy solution, a blanket ban can result in many false positives, which thus results in humans being identified as bots and being banned. Such situations can result in emotional or psychological harm of the human being banned, or toxic online behavior where users repeatedly report another user that they personally dislike as a bot to silence them[112]. Additionally, social media bots do not necessarily work alone: they coordinate with other bots – sometimes even human agents – to push out their agenda[82], and therefore if one agent warrants a ban, should the entire network be banned? To ban an entire network may entangle several unsuspecting humans who have been influenced by the bots to partake in the conversation. With these considerations in mind, regulation is a scope of problem with which to be studied: which types of bots should we ban? What are the activities of a bot that would warrant a ban?

Methods

Examining bot literature

We examined recent bot literature for the definition of "social media bot". For academic definitions, we searched the phrase "social media bot" on Google Scholar. For industry literature, we searched the phrase "social media bot" on Google Search. Then, we manually scanned through the results. We picked out the more relevant and highly cited papers that had a definition of a social media bot. We read through each paper, and manually extracted the definition of a social media bot stated in the paper.

Next, we looked through all the definitions and picked out key phrases. We then harmonized the phrases and definitions to create a general definition of the bot. All authors agreed on the definitions and categorizations.

Data collection and labeling

We collected a dataset from Twitter/X involving global events which provides a richness in a general understanding of the bot and human differentiation. The list of data collection parameters are detailed in Supplementary Material Table S1.

We labeled each user in this dataset as bot or human with the BotHunter algorithm. This algorithm uses a tiered random forest classifier with increasing amounts of user data to evaluate the probability of the user being a bot. The algorithm returns a bot probability score that is between 0 and 1, where scores above 0.7 we deem as a bot, and scores below 0.7 we deem as a human. This 0.7 threshold value is determined from a previous longitudinal study that sought to identify a stable bot score threshold that best represents the automation capacity of a user[71]. This bot algorithm and threshold is chosen so that our studies will be consistent with the original studies of the dataset that used the BotHunter algorithm[9,34,79–82]. We calculated the proportion of bot users against the total number of users within each event. Our results are presented in a bar graph.

Comparison by Psycholinguistic Cues We parse the collected dataset of tweets through the NetMapper software v1.0.82[83] to extract out psycholinguistic cues of the texts. NetMapper extracts the number of each of the cues per tweet. The software returns three types of cues: semantic cues, emotion cues and metadata cues. The linguistic cues are returned by matching words against a dictionary for each category. The dictionary has words in 40 languages. Then, for each user, we average the use of each cue per category, as the trend for the user. We then perform a student t-test comparison between the cues of each user type with Bonferroni correction, and identify whether the cues are significantly different between the bot and human at the $p < 0.05$ level. We then remove the retweets from the Captain Marvel and Black Panther datasets and compare the cue distribution of original tweets with all tweets. This analysis compares the differences in the distribution of cues of tweets originating from the user type and their retweets.

Comparison by Self-Presentation of Identity To classify identities, we compare the user description and bio information against a survey of occupation of United States users performed in 2015[113]. If the occupation is present in the user information, the user is tagged with the identity. A user can have more than one identity. We compare the top identities used by bots and human users across all events. These identities are also divided up into seven categories: religion, race/nationality, political, job, gender, family and others. We then classify each user into these categories of identities. Again, each user can fall into multiple categories.

Next, we examined how different identities frame their posts differently. We examined five topic frames: family, gender, political, race/nationality and religion. We parsed our data through the NetMapper software, which returned us the number of framing cues present in each tweet. These cues are calculated through matching a multilingual lexicon for each frame. Then, for each user, we average the use of each of the topic frames. For each most frequent identity affiliated with by bots and humans, we compare the difference in the average use of each topic frame through a percentage difference calculation. The percentage difference in the use of framing cues is calculated as: $\frac{(H-B)}{H}$, where H is the average use of the framing cue by humans, and B is the average use of framing cue by bots. This comparison tells us how much more bots use a framing cue as compared to humans. If the percentage is negative, bots use the framing cue more than humans. If the percentage is positive, bots use the cue less than humans.

The set of topic frames also corresponds with the identity categories. Therefore, we also compared the identity categories against the average use of each topic frame. This comparison is performed across bots and humans. We plot heatmaps to show the relationship between the average use of each topic frame topic frame against the identity categories.

Comparison by Social Interactions We construct the all-communication ego-networks of the users in our dataset. We analyzed all the users for Asian Elections, Black Panther, Canadian Elections 2019, Captain Marvel and ReOpen America events. Due to the size of the data, we analyzed a 2% sample of users of the

	Bots	Humans
Volume (%)	21.9 ± 9.8	78.1 ± 9.8
Psycholinguistic Cues	Use more hashtags, mentions; has more tweets/hour, total tweets, friends:followers ratio	uses more media, favorites, replies, quotes, urls
Self-presentation of Identity	Concentrate their affiliations on a few identities	Have a more varied identity affiliations
Have identity affiliation (%)	21.4±5.7	27.0±9.2
Topic Frames	Political topics	Family and Gender
Identity vs Topic Frames	Converse about topics that closely match their identity	Have a larger range of topics
Social Interactions	Star communication structure	Tiered communication structure
	Denser interaction networks	Less dense interaction networks
	Interact with more human than bot alters	Interact with more human than bot alters

Table 7. Summary of differences between bots and humans.

Coronavirus2020-2021 users ($N = 4.6$mil), and a 50% sample of users from the US Elections 2020 ($N = 500$ k). The ego-networks are network graphs of the bot and human users in focus. In the networks, each user is represented as a node, and a communication interaction between users are represented as links. The ego-networks are constructed using all-communication interactions, that is any communication between users (i.e., retweet, @mentions, reply, quote) is reflected as a link. We analyzed the network properties of the ego-networks constructed per event. These properties are: total-degree, in degree, out degree, density. We also analyzed the number of bot and human alters there are in the ego networks. No pre-processing was performed on the networks prior to the calculations. We used the ORA software v.3.0.9.181 to load in the networks and perform the calculations[83]. We finally visualize the network graphs of one- and two- degree ego networks of a sample of bots and humans Fig. 6. These are the 20 most frequent communicators in the Asian Elections sub-dataset A 1-degree network shows alters (connected users) that are in direct communication with the user, and a 2-degree network shows alters in direct communication with the 1st-degree alters.

Conclusion

Social media bots are deeply interweaved into our digital ecosystem. More than half of the Internet traffic in 2023 was generated by these AI agents[114]. Bots are able to generate this volume of traffic because of their use of automation, which enables them to create more content and form more relationships. This article surmised a definition of a social media bot based on the three elements that a social media platform contains: user, content, interactions. Our definition breaks down the automation on social media platforms into its core mechanics, and therefore provides the foundation for further research, analysis and policies regulating the digital space. We performed a large scale data analysis of bot and human characteristics across events around the globe, presenting the uniqueness of the bot species from a macro perspective: how bots and humans differ in terms of the use of linguistic cues, social identity affiliations and social interactions. On a global scale, bots and humans do have consistent differences, which can be used to differentiate the two species of users. Table 7 summarizes the differences between bots and humans as a conclusive remark. Finally, we provide recommendations for the use and regulation of bots. These recommendations are informed by our results. We also lay out the challenges and opportunities for the future of bot detection in a "Detect, Differentiate, Disrupt" frame. We invite academics, non-profits , and policymakers to take part in this active research area.

Data availability

Please contact the authors to request the data and code from this study, in accordance to the data sharing policies of the social media platforms.

Received: 1 January 2025; Accepted: 27 March 2025
Published online: 31 March 2025

References

1. Woolley, S. C. Automating power: Social bot interference in global politics. *First Monday* (2016).
2. Lotan, G. et al. The arab spring| the revolutions were tweeted: Information flows during the 2011 tunisian and egyptian revolutions. *Int. J. Commun.* **5**, 31 (2011).
3. Ng, L. H. X., Robertson, D. C. & Carley, K. M. Cyborgs for strategic communication on social media. *Big Data Soc.* **11**, 20539517241231276 (2024).
4. Ng, L. H. X. & Carley, K. M. Assembling a multi-platform ensemble social bot detector with applications to us 2020 elections. *Soc. Netw. Anal. Min.* **14**, 45 (2024).
5. Chang, H.-C. H., Chen, E., Zhang, M., Muric, G. & Ferrara, E. Social bots and social media manipulation in 2020: The year in review. In *Handbook of Computational Social Science, Volume 1*, 304–323 (Routledge, 2021).
6. Seckin, O. C., Atalay, A., Otenen, E., Duygu, U. & Varol, O. Mechanisms driving online vaccine debate during the covid-19 pandemic. *Soc. Media+ Soc.* **10**, 20563051241229656 (2024).
7. Ferrara, E. What types of covid-19 conspiracies are populated by twitter bots? arXiv preprint arXiv:2004.09531 (2020).
8. Ng, L. H. X. & Carley, K. M. Pro or anti? a social influence model of online stance flipping. *IEEE Trans. Network Sci. Eng.* **10**, 3–19 (2022).
9. Magelinski, T., Ng, L. H. X. & Carley, K. M. A synchronized action framework for responsible detection of coordination on social media. arXiv preprint arXiv:2105.07454 (2021).

10. Broniatowski, D. A. et al. Weaponized health communication: Twitter bots and russian trolls amplify the vaccine debate. *Am. J. Public Health* **108**, 1378–1384 (2018).
11. Shao, C. et al. The spread of low-credibility content by social bots. *Nat. Commun.* **9**, 1–9 (2018).
12. Ng, L. H. X., Cruickshank, I. J. & Carley, K. M. Cross-platform information spread during the january 6th capitol riots. *Soc. Netw. Anal. Min.* **12**, 133 (2022).
13. Ingram, M. Musk's Twitter bid, and the 'bot' complication. https://www.cjr.org/the_media_today/musks-twitter-bid-and-the-bot-complication.php (2022). [Accessed 28-10-2024].
14. Childs, J. Elon Musk says X is fighting bots and spam, and the solution is: $1 subscriptions — latimes.com. https://www.latimes.com/business/story/2023-10-18/x-pilot-program-to-charge-1-a-year-in-effort-to-combat-bots-spam (2023). [Accessed 28-10-2024].
15. Ellaky, Z., Benabbou, F. & Ouahabi, S. Systematic literature review of social media bots detection systems. *J. King Saud Univ.-Comput. Inf. Sci.* **35**, 101551 (2023).
16. Ng, L. H. X., Bartulovic, M. & Carley, K. M. Tiny-botbuster: Identifying automated political coordination in digital campaigns. In: *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, 25–34 (Springer, 2024).
17. Beskow, D. M. & Carley, K. M. Bot-hunter: a tiered approach to detecting & characterizing automated activity on twitter. In: *Conference paper. SBP-BRiMS: International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation*, vol. 3 (2018).
18. Sayyadiharikandeh, M., Varol, O., Yang, K.-C., Flammini, A. & Menczer, F. Detection of novel social bots by ensembles of specialized classifiers. In: *Proceedings of the 29th ACM international conference on information & knowledge management*, 2725–2732 (2020).
19. Ng, L. H. X. & Carley, K. M. Botbuster: Multi-platform bot detection using a mixture of experts. in: *Proceedings of the international AAAI conference on web and social media* **17**, 686–697 (2023).
20. Orabi, M., Mouheb, D., Al Aghbari, Z. & Kamel, I. Detection of bots in social media: a systematic review. *Inform. Processing Manage.* **57**, 102250 (2020).
21. Kolomeets, M., Chechulin, A. & Kotenko, I. V. Bot detection by friends graph in social networks. *J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl.* **12**, 141–159 (2021).
22. Li, S. et al. Botfinder: a novel framework for social bots detection in online social networks based on graph embedding and community detection. *World Wide Web* **26**, 1793–1809 (2023).
23. Feng, S. *et al.* What does the bot say? opportunities and risks of large language models in social media bot detection. arXiv preprint arXiv:2402.00371 (2024).
24. Jacobs, C. S., Ng, L. H. X. & Carley, K. M. Tracking china's cross-strait bot networks against taiwan. In: *International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation*, 115–125 (Springer, 2023).
25. Uyheng, J. & Carley, K. M. Bot impacts on public sentiment and community structures: Comparative analysis of three elections in the asia-pacific. In: *Social, Cultural, and Behavioral Modeling: 13th International Conference, SBP-BRiMS 2020, Washington, DC, USA, October 18–21, 2020, Proceedings 13*, 12–22 (Springer, 2020).
26. Bessi, A. & Ferrara, E. Social bots distort the 2016 us presidential election online discussion. *First monday***21** (2016).
27. Khaund, T., Al-Khateeb, S., Tokdemir, S. & Agarwal, N. Analyzing social bots and their coordination during natural disasters. In: *Social, Cultural, and Behavioral Modeling: 11th International Conference, SBP-BRiMS 2018, Washington, DC, USA, July 10-13, 2018, Proceedings 11*, 207–212 (Springer, 2018).
28. Ng, L. H. & Taeihagh, A. How does fake news spread? understanding pathways of disinformation spread through apis. *Policy Internet* **13**, 560–585 (2021).
29. Hajli, N., Saeed, U., Tajvidi, M. & Shirazi, F. Social bots and the spread of disinformation in social media: the challenges of artificial intelligence. *Br. J. Manag.* **33**, 1238–1253 (2022).
30. Kenny, R., Fischhoff, B., Davis, A., Carley, K. M. & Canfield, C. Duped by bots: why some are better than others at detecting fake social media personas. *Hum. Factors* **66**, 88–102 (2024).
31. Kolomeets, M., Tushkanova, O., Desnitsky, V., Vitkova, L. & Chechulin, A. Experimental evaluation: Can humans recognise social media bots?. *Big Data Cognitive Comput.* **8**, 24 (2024).
32. Carley, K. M. Social cybersecurity: an emerging science. *Comput. Math. Organ. Theory* **26**, 365–381 (2020).
33. Ferrara, E., Varol, O., Davis, C., Menczer, F. & Flammini, A. The rise of social bots. *Commun. ACM* **59**, 96–104 (2016).
34. Uyheng, J., Ng, L. H. X. & Carley, K. M. Active, aggressive, but to little avail: characterizing bot activity during the 2020 singaporean elections. *Comput. Math. Organ. Theory* **27**, 324–342 (2021).
35. Himelein-Wachowiak, M. et al. Bots and misinformation spread on social media: Implications for covid-19. *J. Med. Internet Res.* **23**, e26933 (2021).
36. Ng, L. H. X., Zhou, W. & Carley, K. M. Exploring cognitive bias triggers in covid-19 misinformation tweets: A bot vs. human perspective. arXiv preprint arXiv:2406.07293 (2024).
37. Cloudflare. What is a social media bot? | social media bot definition. https://www.cloudflare.com/learning/bots/what-is-a-social-media-bot/. [Accessed 28-10-2024].
38. Assenmacher, D. et al. Demystifying social bots: On the intelligence of automated social media actors. *Soc. Media+ Soc.* **6**, 2056305120939264 (2020).
39. DHS, U. D. o. H. S. niccs.cisa.gov. https://niccs.cisa.gov/sites/default/files/documents/pdf/ncsam_socialmediabotsoverview_508.pdf?trackDocs=ncsam_socialmediabotsoverview_508.pdf (2018). [Accessed 29-10-2024].
40. Danaditya, A., Ng, L. H. X. & Carley, K. M. From curious hashtags to polarized effect: profiling coordinated actions in indonesian twitter discourse. *Soc. Netw. Anal. Min.* **12**, 105 (2022).
41. Hayawi, K., Saha, S., Masud, M. M., Mathew, S. S. & Kaosar, M. Social media bot detection with deep learning methods: a systematic review. *Neural Comput. Appl.* **35**, 8903–8918 (2023).
42. Tsvetkova, M., García-Gavilanes, R., Floridi, L. & Yasseri, T. Even good bots fight: The case of wikipedia. *PLoS One* **12**, e0171774 (2017).
43. Stieglitz, S., Brachten, F., Ross, B. & Jung, A.-K. Do social bots dream of electric sheep? a categorisation of social media bot accounts. arXiv preprint arXiv:1710.04044 (2017).
44. CISA, C. & Agency, I. S. Social media bots. https://www.cisa.gov/sites/default/files/publications/social_media_bots_infographic_set_508.pdf. [Accessed 28-10-2024].
45. Hofeditz, L., Ehnis, C., Bunker, D., Brachten, F. & Stieglitz, S. Meaningful use of social bots? possible applications in crisis communication during disasters. In *ECIS* (2019).
46. Piccolo, L. S. G., Troullinou, P. & Alani, H. Chatbots to support children in coping with online threats: Socio-technical requirements. In: *Designing Interactive Systems Conference 2021*, DIS '21, 1504-1517, https://doi.org/10.1145/3461778.3462114 (Association for Computing Machinery, New York, NY, USA, 2021).
47. Krueger, J. & Osler, L. Communing with the dead online: chatbots, grief, and continuing bonds. *J. Conscious. Stud.* **29**, 222–252 (2022).
48. Boshmaf, Y., Muslukhov, I., Beznosov, K. & Ripeanu, M. The socialbot network: when bots socialize for fame and money. In: *Proceedings of the 27th annual computer security applications conference*, 93–102 (2011).

49. Howard, P. N. & Parks, M. R. Social media and political change: Capacity, constraint, and consequence (2012).
50. Chavoshi, N., Hamooni, H. & Mueen, A. Debot: Twitter bot detection via warped correlation. *In Icdm* **18**, 28–65 (2016).
51. Stieglitz, S. *et al.* Do social bots (still) act different to humans?–comparing metrics of social bots with those of humans. In: *Social Computing and Social Media. Human Behavior: 9th International Conference, SCSM 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9-14, 2017, Proceedings, Part I 9*, 379–395 (Springer, 2017).
52. Grimme, C., Preuss, M., Adam, L. & Trautmann, H. Social bots: Human-like by means of human control?. *Big Data* **5**, 279–293 (2017).
53. Alsmadi, I. & O'Brien, M. J. How many bots in russian troll tweets?. *Inform. Process. Manage.* **57**, 102303 (2020).
54. Ng, L. H. X. & Carley, K. M. Deflating the chinese balloon: types of twitter bots in us-china balloon incident. *EPJ Data Sci.* **12**, 63 (2023).
55. Shao, C., Ciampaglia, G. L., Varol, O., Flammini, A. & Menczer, F. The spread of fake news by social bots. arXiv preprint arXiv:1707.07592**96**, 14 (2017).
56. Vziatysheva, V. How fake news spreads online? *International Journal of Media & Information Literacy***5** (2020).
57. Lokot, T. & Diakopoulos, N. News bots: Automating news and information dissemination on twitter. *Digit. J.* **4**, 682–699 (2016).
58. Gorwa, R. & Guilbeault, D. Unpacking the social media bot: A typology to guide research and policy. *Policy Internet* **12**, 225–248 (2020).
59. Arora, A., Arora, A. & McIntyre, J. Developing chatbots for cyber security: Assessing threats through sentiment analysis on social media. *Sustainability* **15**, 13178 (2023).
60. Liu, L., Preotiuc-Pietro, D., Samani, Z. R., Moghaddam, M. E. & Ungar, L. Analyzing personality through social media profile picture choice. In: *Proceedings of the International AAAI Conference on Web and Social Media* **10**, 211–220 (2016).
61. Ng, L. H. X. & Cruickshank, I. J. Recruitment promotion via twitter: a network-centric approach of analyzing community engagement using social identity. *Digital Govern. Res. Practice* **4**, 1–17 (2023).
62. McKelvey, F. & Dubois, E. Computational propaganda in canada: The use of political bots (2017).
63. Woolley, S. C. & Howard, P. N. *Computational propaganda: Political parties, politicians, and political manipulation on social media* (Oxford University Press, 2018).
64. Zouzou, Y. & Varol, O. Unsupervised detection of coordinated fake-follower campaigns on social media. *EPJ Data Sci.* **13**, 62 (2024).
65. Albadi, N., Kurdi, M. & Mishra, S. Hateful people or hateful bots? detection and characterization of bots spreading religious hatred in arabic social media. In: *Proceedings of the ACM on Human-Computer Interaction* **3**, 1–25 (2019).
66. Blane, J. T. *Social-Cyber Maneuvers for Analyzing Online Influence Operations*. Ph.D. thesis, United States Military Academy (2023).
67. Badawy, A., Ferrara, E. & Lerman, K. Analyzing the digital traces of political manipulation: The 2016 russian interference twitter campaign. In: *2018 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)*, 258–265 (IEEE, 2018).
68. Schwartz, O. In 2016, microsoft's racist chatbot revealed the dangers of online conversation. *IEEE Spectr.* **11**, 2019 (2019).
69. Brachten, F. *et al.* Threat or opportunity?-examining social bots in social media crisis communication. arXiv preprint arXiv:1810.09159 (2018).
70. Martini, F., Samula, P., Keller, T. R. & Klinger, U. Bot, or not? Comparing three methods for detecting social bots in five political discourses. *Big Data Soc.* **8**, 20539517211033570 (2021).
71. Ng, L. H. X., Robertson, D. C. & Carley, K. M. Stabilizing a supervised bot detection algorithm: How much data is needed for consistent predictions?. *Online Soc. Networks Media* **28**, 100198 (2022).
72. Tan, Z. *et al.* BotPercent: Estimating bot populations in Twitter communities. In Bouamor, H., Pino, J. & Bali, K. (eds.) *Findings of the Association for Computational Linguistics: EMNLP 2023*, 14295–14312, https://doi.org/10.18653/v1/2023.findings-emnlp.954 (Association for Computational Linguistics, Singapore, 2023).
73. Yan, H. Y., Yang, K.-C., Shanahan, J. & Menczer, F. Exposure to social bots amplifies perceptual biases and regulation propensity. *Sci. Rep.* **13**, 20707 (2023).
74. Yan, H. Y. & Yang, K.-C. The landscape of social bot research: A critical appraisal. In *Handbook of Critical Studies of Artificial Intelligence*, 716–725 (Edward Elgar Publishing, 2023).
75. Yang, K.-C. & Menczer, F. Anatomy of an ai-powered malicious social botnet. *Journal of Quantitative Description: Digital Media***4** (2024).
76. Microsoft. How to spot bots on social media - Microsoft 365 — microsoft.com. https://www.microsoft.com/en-us/microsoft-365-life-hacks/privacy-and-safety/how-to-spot-bots-on-social-media (2024). [Accessed 29-10-2024].
77. Simpson, M. Social Media Bots 101 - All You Need to Know — meltwater.com. https://www.meltwater.com/en/blog/social-media-bots (2024). [Accessed 29-10-2024].
78. Imperva. What are Bots | Bot Types & Mitigation Techniques | Imperva — imperva.com. https://www.imperva.com/learn/application-security/what-are-bots/. [Accessed 29-10-2024].
79. Babcock, M., Beskow, D. M. & Carley, K. M. Beaten up on twitter? exploring fake news and satirical responses during the black panther movie event. In: *Social, Cultural, and Behavioral Modeling: 11th International Conference, SBP-BRiMS 2018, Washington, DC, USA, July 10-13, 2018, Proceedings 11*, 97–103 (Springer, 2018).
80. King, C., Bellutta, D. & Carley, K. M. Lying about lying on social media: a case study of the 2019 canadian elections. In: *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, 75–85 (Springer, 2020).
81. Babcock, M., Villa-Cox, R. & Carley, K. M. Pretending positive, pushing false: Comparing captain marvel misinformation campaigns. *Disinformation, Misinformation, and Fake News in Social Media: Emerging Research Challenges and Opportunities* 83–94 (2020).
82. Ng, L. H. X. & Carley, K. M. A combined synchronization index for evaluating collective action social media. *Appl. Network Sci.* **8**, 1 (2023).
83. Carley, L. R., Reminga, J. & Carley, K. M. Ora & netmapper. In *International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation. Springer*, 1, 2–1 (2018).
84. Stella, M., Ferrara, E. & De Domenico, M. Bots increase exposure to negative and inflammatory content in online social systems. *Proc. Natl. Acad. Sci.* **115**, 12435–12440 (2018).
85. Aldayel, A. & Magdy, W. Characterizing the role of bots' in polarized stance on social media. *Soc. Netw. Anal. Min.* **12**, 30 (2022).
86. Ferrara, E. Disinformation and social bot operations in the run up to the 2017 french presidential election. arXiv preprint arXiv:1707.00086 (2017).
87. Pathak, A., Madani, N. & Joseph, K. A method to analyze multiple social identities in twitter bios. In: *Proceedings of the ACM on Human-Computer Interaction* **5**, 1–35 (2021).
88. Van der Walt, E., Eloff, J. H. & Grobler, J. Cyber-security: Identity deception detection on social media platforms. *Comput. Security* **78**, 76–89 (2018).
89. Radivojevic, K., Clark, N. & Brenner, P. Llms among us: Generative ai participating in digital discourse. In: *Proceedings of the AAAI Symposium Series* **3**, 209–218 (2024).
90. Howard, P. N. *Lie machines: How to save democracy from troll armies, deceitful robots, junk news operations, and political operatives* (Yale University Press, 2020).

91. Tardelli, S., Avvenuti, M., Tesconi, M. & Cresci, S. Characterizing social bots spreading financial disinformation. In: *International conference on human-computer interaction*, 376–392 (Springer, 2020).
92. DeVerna, M. R., Yan, H. Y., Yang, K.-C. & Menczer, F. Artificial intelligence is ineffective and potentially harmful for fact checking. arXiv preprint arXiv:2308.10800 (2023).
93. Hill, J., Ford, W. R. & Farreras, I. G. Real conversations with artificial intelligence: A comparison between human-human online conversations and human-chatbot conversations. *Comput. Hum. Behav.* **49**, 245–250 (2015).
94. Bisgin, H., Agarwal, N. & Xu, X. A study of homophily on social media. *World Wide Web* **15**, 213–232 (2012).
95. Ng, L. H. X. & Carley, K. M. Online coordination: methods and comparative case studies of coordinated groups across four events in the united states. In: *Proceedings of the 14th ACM Web Science Conference 2022*, 12–21 (2022).
96. Mazza, M., Cresci, S., Avvenuti, M., Quattrociocchi, W. & Tesconi, M. Rtbust: Exploiting temporal patterns for botnet detection on twitter. In: *Proceedings of the 10th ACM conference on web science*, 183–192 (2019).
97. Russell, S. J. & Norvig, P. *Artificial intelligence: A modern approach* (Pearson, 2016).
98. Lu, H.-C. & Lee, H.-w. Agents of discord: Modeling the impact of political bots on opinion polarization in social networks. *Social Science Computer Review* 08944393241270382 (2024).
99. Chen, Y. The social influence of bots and trolls in social media. In: *Handbook of Computational Social Science, Volume 1*, 287–303 (Routledge, 2021).
100. Ng, L. H. X., Kloo, I., Clark, S. & Carley, K. M. An exploratory analysis of covid bot vs human disinformation dissemination stemming from the disinformation dozen on telegram. *Journal of Computational Social Science* 1–26 (2024).
101. Jacobs, C. S., Ng, L. H. X. & Carley, K. M. Detection evasion techniques of state-sponsored accounts. In: *Proceedings of the ICWSM Workshops*, https://doi.org/10.36190/2024.63 (Buffalo, New York, USA, 2024).
102. Tyagi, A., Babcock, M., Carley, K. M. & Sicker, D. C. Polarizing tweets on climate change. In: *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, 107–117 (Springer, 2020).
103. Yang, K.-C., Ferrara, E. & Menczer, F. Botometer 101: Social bot practicum for computational social scientists. *J. Comput. Soc. Sci.* **5**, 1511–1528 (2022).
104. Tardelli, S., Avvenuti, M., Tesconi, M. & Cresci, S. Detecting inorganic financial campaigns on twitter. *Inf. Syst.* **103**, 101769 (2022).
105. Marigliano, R., Ng, L. H. X. & Carley, K. M. Analyzing digital propaganda and conflict rhetoric: a study on russia's bot-driven campaigns and counter-narratives during the ukraine crisis. *Soc. Netw. Anal. Min.* **14**, 170 (2024).
106. Ng, L. H. X., Cruickshank, I. J. & Farr, D. Building bridges between users and content across multiple platforms during natural disasters. arXiv preprint arXiv:2502.02681 (2025).
107. Schroeder, R. Big data and the brave new world of social media research. *Big Data Soc.* **1**, 2053951714563194 (2014).
108. Gillespie, T. *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media* (Yale University Press, 2018).
109. Cresci, S. Detecting malicious social bots: story of a never-ending clash. In: *Multidisciplinary International Symposium on Disinformation in Open Online Media*, 77–88 (Springer, 2020).
110. Burgess, M. Millions of People Are Using Abusive AI 'Nudify' Bots on Telegram — wired.com. https://www.wired.com/story/ai-deepfake-nudify-bots-telegram/ (2024). [Accessed 21-10-2024].
111. Ferrara, E. Contagion dynamics of extremist propaganda in social networks. *Inf. Sci.* **418**, 1–12 (2017).
112. Jones, M. L. Silencing bad bots: Global, legal and political questions for mean machine communication. *Commun. Law Policy* **23**, 159–195 (2018).
113. Smith-Lovin, L. *et al.* Mean affective ratings of 929 identities, 814 behaviors, and 660 modifiers by university of georgia and duke university undergraduates and by community members in durham, nc, in 2012-2014. *University of Georgia: Distributed at UGA Affect Control Theory Website:* http://research.franklin.uga.edu/act (2016).
114. LaFrance, A. The Internet Is Mostly Bots — theatlantic.com. https://www.theatlantic.com/technology/archive/2017/01/bots-bots-bots/515043/ (2017). [Accessed 27-09-2024].

Acknowledgements

The first author thanks Christine Lepird for her ideas and comments towards this work. This material is based upon work supported by the Scalable Technologies for Social Cybersecurity, U.S. Army (W911NF20D0002), the Minerva-Multi-Level Models of Covert Online Information Campaigns, Office of Naval Research (N000142112765), the Threat Assessment Techniques for Digital Data, Office of Naval Research (N000142412414), and the MURI: Persuasion, Identity & Morality in Social-Cyber Environments (N000142112749), Office of Naval Research. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing official policies, either expressed or implied by the Office of Naval Research, U.S. Army or the U.S. government.

Author contributions

L.H.X.N. conceived the experiments, collected the data and performed the analysis. K.M.C. reviewed the analysis. All authors reviewed the manuscript.

Declarations

Competing interest

The authors declare that there is no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-96372-1.

Correspondence and requests for materials should be addressed to L.H.X.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.