INTERNET POLICY REVIEW
Journal on internet regulation

Volume 12 **Issue** 1

RESEARCH ARTICLE

OPEN ACCESS

PEER REVIEWED

Fake accounts on social media, epistemic uncertainty and the need for an independent auditing of accounts

Martin Moore *King's College London*

DOI: https://doi.org/10.14763/2023.1.1680

Published: 7 February 2023
Received: 8 September 2021 **Accepted:** 18 October 2021

Competing Interests: The author has declared that no competing interests exist that have influenced the text.

Citation: Moore, M. (2023). Fake accounts on social media, epistemic uncertainty and the need for an independent auditing of accounts. *Internet Policy Review*, *12*(1). https://doi.org/10.14763/2023.1.1680

Keywords: Fake accounts, Social media, Digital platforms, Platform policies

Abstract: This article seeks to understand why so little is known about the nature and extent of 'fake or spam accounts' across the leading social media companies, why this lack of knowledge is even greater outside the companies themselves and the implications of this epistemic uncertainty. It concludes that authorities, investors and the public should not be solely reliant on the companies for user account figures, and that there is a need for regular, independent, external audits of inauthentic accounts on social media.

Introduction

On May 13, 2022, Elon Musk suspended his planned purchase of Twitter Inc. 'pending details supporting calculation that spam/fake accounts do indeed represent less than 5% of users' (Musk, 2022a). His tweet included a link to a news article on the recent Twitter filing, in which the company estimated spam and fake accounts comprised less than 5% of users. Almost two months later, on July 8, lawyers acting for Musk wrote to Twitter saying that Musk was terminating his offer to buy the company, claiming that despite his efforts, Twitter had 'failed or refused to provide' sufficient information about false accounts for him to make an adequate assessment of their prevalence (Ringler, 2022).

Critics of Musk's deal quickly wrote off this rationale as an excuse to renegotiate his original offer or find a way out (Fontanella-Khan & Murphy, 2022). Whether or not this is true, Musk's apprehensiveness reflected widespread and long held concerns about the proportion of Twitter profiles that are not authentic users (Ferrara et al., 2016). Nor are these concerns limited to Twitter. All the major social media platforms acknowledge that a proportion of their user bases are bots, spam or false accounts. When these companies were young these inauthentic accounts were considered relatively inconsequential. Now that they are among the most valuable and influential companies in the world, the European Commission terms them the 'gatekeepers' of the virtual world and new laws seek to minimise the harms they cause, there is a need to reconsider the importance of these false accounts (Digital Markets Act, 2020; Online Safety Bill, 2022). Not only is this information, as Elon Musk's lawyers wrote, fundamental to the 'business and financial performance' of social media companies like Twitter, it is also core to the prevailing business model of the internet (digital advertising), and these inauthentic accounts are purported to be the source of co-ordinated abuse, amplification and influence operations (Twitter Safety, 2020; Gleicher et al., 2021). Yet, as Twitter's responses to the calls made by Elon Musk showed, even the companies themselves do not appear to know their full nature or the extent to which they are prevalent.

This article seeks to understand why so little is known about the nature and extent of these 'false or spam accounts' across the leading social media companies, why this lack of knowledge is even greater outside the companies themselves and the implications of this epistemic uncertainty. It concludes that authorities, investors and the public should not be solely reliant on the companies for user account figures, and that there is a need for regular, independent, external audits of inauthentic accounts on social media. The first section of the article considers how much major social media companies appear to know about the number of fake ac-

counts on their services (based on public statements and reporting). The next section assesses why social media companies themselves do not appear to know the full extent of false accounts on their services. The article then examines why external observers have found it harder still to assess the nature and scale of different profile types, before considering the consequences of fake accounts and of persistent epistemic uncertainty. It concludes by outlining the need for future auditing and research, noting the relative absence of any requirements for the auditing of user accounts within current platform policy initiatives.

It is important to note that despite the normative implications of the terminology, 'false', 'fake', 'spam' or 'bot' profiles on social media are not necessarily a problem, and in many instances may provide positive benefits. Having multiple social media profiles allows people to avoid 'context collapse' and take on different roles or personas so they can perform different functions in different aspects of their lives (Marwick & boyd, 2011). Many social bots perform useful and productive tasks, and have the potential to do more (e.g. Hofeditz et al., 2019). Yet, as this article will illustrate, the number of these profiles and the epistemic uncertainty about them, creates political, social and economic problems, which if not addressed, will make the resolution of other negative externalities associated with social media even more difficult.

Methodologically, the article adopts a mixed methods approach, necessary due to the multifarious sources of information about social media profiles, and because of the extensive gaps in our current knowledge. It benefits from extensive computer science literature, particularly on bots, though it does not employ computer science methods. Rather, it uses evidence produced by the major social media companies themselves (including corporate reports, statements, filings and public posts), supplemented by industry surveys, journalistic investigations and internet archival research (via archive.org and the Zuckerberg Files). To contextualise the research and better understand the perspectives of the social media companies, a series of 30-60 minute, targeted, in-person interviews were conducted – mostly via video conference – with the relevant representatives from Twitter and Meta, as well as with researchers who have sought to identify specific categories of inauthentic social media users and the uses to which these profiles are employed (see Appendix for details). This research is contextualised when appropriate. In addition, a systematic literature review was conducted, understood as "a systematic, explicit, and reproducible method for identifying, evaluating, and synthesizing the existing body of completed and recorded work produced by researchers, scholars, and practitioners" (Fink, 2010) to identify the research to date on inauthentic social media

users.

The article employs terminology used by the companies themselves, though recognises that this terminology can be both ambiguous and can encourage misleading normative assumptions. Calling accounts 'fake', 'false' or 'spam', as the companies do, makes them appear deceptive and perfidious, even though their definition was originally driven by economic, rather than ethical considerations (to distinguish monetizable from non-monetizable users). In order to group these terms together, and again build on the language used by the companies, this article will refer to *inauthentic* accounts or profiles, while again recognizing the insufficiency of this term due, in part, to epistemic uncertainty about the nature of these accounts.

Estimated proportion of inauthentic accounts on leading social media services

Since the leading social media companies have gone public and been required to report on their users, they have acknowledged that a proportion of their users are not directly associated with a single individual person. In 2012, Facebook[1] admitted that around 83 million of their active users – equivalent to the population of Germany – were 'duplicate' or 'false' (Facebook Inc., 2012). After 2012 the proportion of duplicate and false users on the service rose, according to Facebook, from 7-9% of users to 13-14% by 2017 (Facebook Inc., 2017). Given how quickly the absolute number of Facebook users also grew in the period, this equated to about 260 million duplicate and false profiles in 2017. The following year Facebook reported that the number of these users rose again, to approximately 16%, where it remained in 2019 and 2020. In 2020, therefore, Facebook reported that 448 million accounts on its platform were duplicate or false. It did not record or report the figure for Instagram, whose users were reported to have reached one billion in 2018 (Stern, personal communication, 25 June, 2021).

Twitter also provides approximate figures for the proportion of false, spam or automated accounts in its Securities filings. In its 2013 Annual Report the company 'estimated that false or spam accounts represented less than 5% of our MAUs' (Twitter Inc., 2013, p. 2). This figure was differentiated from 11% of active users who use 'applications that automatically contact our servers for regular updates with no user action involved' (Twitter Inc., 2013, p. 2). It was not made clear whether these

1. This article examines social media services and therefore refers to each by name (i.e. Facebook, Instagram, WhatsApp) rather than by the name of the parent company (Facebook rebranded as Meta in October 2021). Meta is used only when referring to the parent company.

proportions overlapped or were complementary, nor whether the second figure applied exclusively to automated accounts. The following year the annual report repeated the 5% figure, with the same caveats, though reduced the 11% to 8.5% (Twitter Inc., 2014, p. 4). The 5% and 8.5% figures were then repeated, word-for-word, in the annual reports for fiscal years ending December 2015, 2016 and 2017 (Twitter Inc., 2016; 2017; 2018). Twitter stopped reporting the second figure from 2018, though continued to repeat the 5% figure (now for mDAU). Other social media services, such as Snap and LinkedIn, also admit to having duplicate, multiple or inauthentic users on their platforms, though do not provide numbers for them (Snap Inc., 2021a, p. 4; LinkedIn Inc., 2016, p. 18).

Each of the leading global social media services, therefore, admits that a proportion of the profiles on its service are spam, fake or duplicate. Though, as will be illustrated below, the actual number of these inauthentic accounts, and what makes them inauthentic, remains far from clear.

Acknowledgement of epistemic uncertainty

The companies partially admit their inability to identify inauthentic users, and that their figures are guesstimates that could be wildly off. Each year since Facebook began reporting publicly on its users in 2012, it has set out the same caveat: 'there are inherent challenges in measuring usage of our products across large online and mobile populations around the world' (Facebook Inc., 2012-2020). Every year Facebook explains that it estimated the number of duplicate or false users based on 'an internal review of a limited sample of accounts', and that it applied 'significant judgment in making this determination' (Facebook Inc., 2012-2020). 'Duplicate and false accounts', the company acknowledged in 2018, 'are very difficult to measure at our scale, and it is possible that the actual number of duplicate and false accounts may vary significantly from our estimates' (Facebook Inc., 2018, p. 4). In 2020, after Facebook introduced its 'Family metrics' to measure users across Facebook, Instagram and WhatsApp, it extended these caveats further, saying that 'it is very difficult to attribute multiple user accounts within and across products to individual people, and it is possible that the actual numbers of unique people using our products may vary significantly from our estimates, potentially beyond our estimated error margins' (Facebook Inc., 2020, p. 5).

Twitter employs almost the same language as Facebook when providing warnings around its user figures. 'There are', it has said since 2013, 'inherent challenges in measuring usage and user engagement across our large user base around the world', and 'In making this determination [as to the proportion of inauthentic

users], we applied significant judgment, so our estimation of false or spam accounts may not accurately represent the actual number of such accounts, and the actual number of false or spam accounts could be higher than we have estimated' (Twitter Inc., 2013, p. 30). This uncertainty may be due to the method used to calculate inauthentic accounts. According to Elon Musk, Twitter uses a random sample of only 100 Twitter profiles to calculate the proportion of fake/spam/duplicate accounts (Musk, 2022b).

Other leading US social media services do not even try to estimate the actual number of inauthentic accounts on their service: 'there may', Snap stated, 'be individuals who have unauthorized or multiple Snapchat accounts… We have not determined the number of such multiple accounts' (Snap Inc., 2021a, p. 4). Instagram provides no information about the number of inauthentic users. This paper identifies three main reasons for this epistemic uncertainty: inconsistent definition of 'user' and 'false/fake', limited user verification and incomplete or irreconcilable calculations.

Causes of epistemic uncertainty

Inconsistent definition of 'user' and 'false/fake'

The political economy of a social media service is driven by its 'users'. Users are the metric that defines social media companies' performance for investors, that generate their income from advertisers, that determine the content and connections they promote and drive the future development of their services. As the Facebook 2020 annual report states: 'The size of our user base and our users' level of engagement are critical to our success' (Facebook Inc., 2020, p. 13). Hence why, since they have gone public, each of the leading social media companies have explained how they calculate and enumerate their users at the start of annual reports. 'The numbers of our monthly active users (MAUs), daily active users (DAUs), mobile MAUs, and average revenue per user (ARPU)' Facebook set out in its first report in 2012, 'are calculated using internal company data based on the activity of user accounts' (Facebook Inc., 2012, p. 4). Similarly, Twitter set out an analogous approach to its 'users' the following year, and explained how fundamental these metrics are to its business: 'We review a number of metrics, including monthly active users, or MAUs, timeline views, timeline views per MAU and advertising revenue per timeline view, to evaluate our business, measure our performance, identify trends affecting our business, formulate business plans and make strategic decisions' (Twitter Inc., 2013, p. 2). 'User' is the lingua franca of social media companies. Facebook employs the term 'user' 339 times, on average, in each of its annual reports be-

tween 2012 and 2020.

Despite the cross-industry reliance on the term 'user', there is an inconsistency, both across and within social media companies, about the definition of the term. For Facebook, users are 'unique people' and 'the individual people behind user accounts' who must use 'the same name that you use in everyday life' and 'Provide accurate information about yourself' (Facebook Inc., 2020). A duplicate account, an account created for a 'non-human entity', or one that is being used in ways that violate terms of service, all represent invalid Facebook users. At Facebook's sister service Instagram, however, the definition of a valid user is broader. Although its Terms of Service state that 'you' cannot impersonate someone on Instagram, or create an automated account, you do not have to give your actual name (Instagram, 2021). Similarly, at Snapchat, 'you' are (by implication, if not explicitly) a person, though you do not have to use your real name or provide accurate information, and you will not - as at Instagram – 'create more than one account for yourself' (Snap Inc., 2021b). By contrast, Twitter allows a single account holder to have multiple accounts, allows automated accounts, and accounts held by non-human entities.

The lack of a consistent industry definition of what represents a social media 'user' complicates the task of differentiating between types of account, and identifying which accounts should be considered authentic or inauthentic. This does not mean that such a definition is impossible, or that user types cannot be identified. Various academic studies have pointed to the ambiguity of terms like *user*, and called for universal measures of user behaviour (Obar & Wildman, 2015; Trifiro & Gerson, 2019). Other research has differentiated between active and inactive users – or so-called 'lurkers', in order to distinguish between the accounts of those who simply observe rather than post (Tagarelli & Interdonato, 2013).

On the basis of a review of existing literature, this paper proposes a basic typology of social media profiles composed of four types: *alter egos* (manually controlled accounts operated by the person with which they are associated – by either a real name or pseudonym), *social bots* ('a computer algorithm that automatically produces content and interacts with humans on social media', Ferrara et al., 2016), *sybils* (manually controlled accounts deliberately dissociated from the person operating them – sometimes referred to as sockpuppets or catfish), and *digital ghosts* (the social media profiles that persist online after people have died, and are sometimes memorialised or reactivated for posthumous use). Most of the focus to date has been on social bots (a term sometimes used indiscriminately with fake accounts), even though these represent a smaller proportion than the other cate-

gories. According to one recent study, for example, the number of digital ghosts on Facebook is already voluminous and will likely outnumber the living within fifty years (Öhman & Watson, 2019). For the purposes of this paper, any profile that is not manually operated by the person directly associated with that profile — an *alter ego* in the typology above — will be considered inauthentic.

Limited user verification

Since social media companies were established they have required minimal information or verification. The most successful early social media services, Friendster (2003), MySpace (2003), Bebo (2005) and LinkedIn (2002), all simply asked for a name and an email address, as can been seen from their early registration pages (Appendix). Facebook (2004) initially distinguished itself by restricting membership to US university students, starting with its founder's own, Harvard. At that point, Mark Zuckerberg said, 'it's just immediately obvious to you' when a profile is fake (Stanford University, 2005). Yet, when Facebook opened to the public in 2006, despite keeping to its real name policy, it only required a name, an email address, a password and a date of birth, in addition to a captcha to screen bots (Appendix). Twitter (2006) initially asked users to register with a mobile phone number, but quickly dropped this and adopted what had, by late 2006, become the norm in social media – registration simply via name and email (Appendix). The norm was driven by the competitive need to sign up as many users as possible, as Steven Levy wrote of Facebook's Open Reg, 'expanding to everyone *was* the mission' (Levy, 2020, p. 120).

As a consequence of these low hurdles for registration and verification, many of the accounts created in the early years were not directly associated with an individual living person (for example duplicate profiles) or to a human being at all (organisations, pets or bots). Within four days of Twitter launching, for example, developers had already started openly publishing code for the creation of Twitter bots, and in the eight years from 2008-16 the number of Twitter bot repositories on just one publicly available developer website — GitHub — grew to over 4,000 (Kollanyi, 2016). 'Fakesters' were an integral part of Friendster from its inception, and MySpace was found to suffer from 'the rampant use of false profiles' (boyd, 2004; Gehl, 2012). Even Facebook, despite its real name policy, was accused of having, by 2008, 'a nation's worth of fake accounts' (Lanier, 2008). Although the hurdles for account creation have risen, it is still straightforward to create a new profile.

Incomplete or irreconcilable calculations

Although the major social media companies release figures about the number of false or fake accounts, the way these figures are calculated does not identify all the potential false or fake profiles. Similarly, the extent of inauthentic profiles reported in the annual filings does not correlate with other reported figures or activities.

Since Q4 2017 Facebook has been releasing data on 'fake accounts actioned' each quarter. The average for the 14 quarters up to Q1 2021 was 1.3 billion (this does not include fake accounts blocked immediately after creation) and a total of 17.6 billion, or over six times the total number of monthly Facebook users, created in 3.5 years (Facebook Inc., 2021). Of these, Facebook claimed that it identified, on average, 99.5% before they were flagged by users themselves (Facebook Inc., 2021). However, as Facebook acknowledged during personal communication for this article, this figure is simply an addition of the fake accounts identified by Facebook with the fake accounts flagged by users (Stern, personal communication, 25 June, 2021). It does not show the number of 'fake accounts' that Facebook has failed to identify, or that its users have not flagged. Nor does it give any breakdown of the type of 'fake accounts' created or their impact. In 2019 Facebook acknowledged that its reporting of fake accounts was not very illuminating, and that it was 'evaluating if there is a better way to report on fake accounts in future' (Schultz, 2019).

Twitter has a similar issue with false account creation, if on a lesser absolute scale. In May 2018, the company reported they were identifying 'more than 9.9 million potentially spammy or automated accounts per week' (Roth & Harvey, 2018). Yet, despite the identification of over four times as many spam or automated accounts that year (515m) as Twitter's daily active user base (122m), and not accounting for additional potential inauthentic accounts, the company reported that the proportion of false or spam accounts remained exactly the same, at below 5% year-on-year. Similarly, the proportion of false or spam accounts that are reported does not change following Twitter's periodic largescale culls of false or spam accounts. In the nine months preceding May 2018, for example, it challenged more than 50 million accounts, ranging from around 2.5 million some months, to 10 million in others (Roth & Harvey, 2018). In 2020 Twitter responded to growing concerns about the influence of bots on its platform, accepting that there was general confusion about their extent and effect, for which Twitter itself was partly responsible. However, rather than seeking to clarify their extent, Twitter asserted that identifying bots was complex — 'It's not just a binary question of bot or not' — and that it

was better to focus on behaviour, and the 'malicious use of automation' (Roth & Pickles, 2020). While this approach may be pragmatic, it shifts the burden of responsibility from the identification of inauthentic profiles to malevolent behaviour on the platform, making any action necessarily reactive.

External research on social media bots

Outside these companies, external researchers have sought to identify, characterise and enumerate certain categories of inauthentic user, most notably automated accounts on Twitter (Oentaryo et al., 2016; Romanov et al., 2017; Gorwa & Guilbeault, 2020; Orabi et al., 2020). Using supervised machine learning techniques and an extensive training data set, in 2017 Varol et al. classified 'a sample of millions of English-speaking active [Twitter] users' and found that 9-15% were likely to be bots: a figure significantly higher than Twitter's annual estimate of below 5% (Varol et al., 2017). Other studies that have sought to quantify the extent and impact of Twitter bots have focused on particular issues or events, and found that during any one political event, there can be tens of thousands of active bots (Thomas et al., 2012, Luceri et al., 2019; Bastos & Mercea, 2019), and there are instances — such as the US 2016 election — where many more have been estimated (Bessi & Ferrara, 2016).

Separate research shows that most social bots do not focus overtly on political issues or events. When Keller and Klinger analysed the increase in bot accounts during the 2017 German electoral campaigns, from 7.1% to 9.9%, they found that only between 1-2% of the increase could be defined as political (Keller & Klinger, 2019). Bence Kollanyi's analysis of thousands of Twitter bot repositories on GitHub between 2008 and 2016 found only 1% made direct reference to politics (Kollanyi, 2016). Other research suggests that there may be vast numbers of bots tweeting about non-political issues. Echeverria and Zhou came across a Star Wars bot network, 'consisting of more than 350,000 bots tweeting random quotations exclusively from Star Wars novels' (Echeverria & Zhou, 2017). A network of prayer bots, virtually unknown outside the Middle East before 2017, far surpasses this number. A single Islamic prayer app in the network of bots (Du3a.org) accounted for approximately 1.9 million tweets per day, greatly exceeding the estimated number of US election tweets posted by bots in the lead up to November 2016 (Öhman et al., 2019). These glimpses into huge Twitter bot networks that exist far beyond the limits of political debate is indicative of how little we know from outside observation about their total number and proportion.

Even extensive research on Twitter bots demonstrates how challenging it is to

identify bots accurately, comprehensively and consistently. This is in part due to the 'ground truth' dilemma (Gorwa & Guilbeault, 2020; Yang et al., 2019). A researcher cannot be sure they have identified the right bots, or all bots, if there is no master list defining which Twitter accounts are real users, and which are not. Even if a researcher discovers a network of a thousand Twitter accounts they think are bots, they may not be able to verify they are bots, nor confirm they are the only bots out there. In 2015 a DARPA Twitter Bot Challenge, six teams successfully identified pro-vaccination bots from a set of 7,000 Twitter accounts (Subrahmanian et al., 2016; Subrahmanian, personal communication, 9 June, 2020). However, the teams were analysing a controlled data set of 7,000 accounts, not the 320 million monthly active users on Twitter. Not only are social media platforms much bigger, but they are constantly evolving (Luceri et al., 2019). A Twitter user might start as a human and then become a bot. Or a human user might supplement their own tweets with automated tweets. Or a human might tweet like a bot (Gilani et al., 2017). Similarly, any two bot developers will create bots that function differently. Finding criteria that successfully identify certain bots (such as number of tweets per day) may fail to identify other bots. Moreover, as an outside observer, it is only possible to analyse Twitter accounts by their activity. Unless they tweet, they are essentially invisible. There could be millions of dormant Twitter accounts – sleeper profiles – just waiting to be activated for a purpose as yet unknown, as was discovered in Russia in 2011, and the US in 2018 (Thomas et al., 2012; Takacs & McCulloh, 2019).

Limited research on inauthentic accounts beyond Twitter

External research on inauthentic accounts has necessarily been constrained by the amount of information made available by the social media companies (Driscoll & Walker, 2014). Since low-bandwidth text-based information is easily and publicly available from Twitter, in contrast with Facebook and YouTube, for example, Twitter has by far attracted the most academic attention. This is borne out by a systematic literature review conducted for this article. 414 relevant academic journal articles, conference articles and chapters dating from 2009 to 2020 were identified based on keyword searches of SCOPUS, ProQuest and Google scholar, using the terms 'social media', 'bots', 'fake accounts', 'false accounts' or 'inauthentic' accounts. Of these 414, 333 sought to identify, characterise, analyse or measure the impact of social media bots or false accounts (as opposed to testing chatbots or identifying botnets not particular to social media). Of these 333 articles, 77% (258) were wholly or primarily based on Twitter. Of the remainder, 15 articles (4.5%) analysed automated

or false profiles on Facebook or Instagram. Although the focus on Twitter is understandable given data limitations, it means most research is focused on a service that is less than a tenth of the scale of Facebook. Moreover, the focus has not only been on Twitter, but on Twitter bots, as opposed to non-automated accounts. Of the 258 articles focused on Twitter, 64% were about Twitter bots.

The limited research that has been done on the authenticity of profiles beyond Twitter, within and outside the academia, has found evidence to suggest that the figures the companies publish may be inaccurate, and that the number of bought, stolen and manufactured profiles is extensive. In 2018, for example, a number of advertisers started a class action suit against Facebook, asserting that they had found that in certain regions and amongst certain demographics, the number of Facebook profiles far exceeded the actual number of people (based on census numbers), sometimes by a factor of 400% (Singer and Project Therapy v Facebook Inc., 2018). Separately, external researchers have discovered a thriving open market in social media accounts, some of which are stolen, others which are manufactured for sale (Bay & Fredheim, 2019). Studying the 2019 market in social media manipulation, researchers working with NATO StratCom were struck not only by the scale and sophistication of the market in profiles, but by its openness (Bay and Fredheim, personal communication, 11 May, 2020). If these profiles were not inauthentic prior to their sale, then they were subsequent to their sale (it should be noted that the StratCom research was conducted for the purposes of informing the strategic communications capabilities of NATO).

Moreover, the figures published by the major social media platforms do not include sybils or digital ghosts, each of which may be extensive (Öhman & Watson, 2019). Therefore, from the limited internal and external figures that exist, there appear to be significantly more profiles on each major social media service than actual people on the services. However, rather than calling these services overpopulated, with its normative implications, this additional social media population would better be termed 'hyper-population'. In this context the term 'hyper' has three meanings that are distinct from 'over' -population. It indicates that this population exists not (just) in physical space, but in cyberspace, corresponding to Jean Baudrillard's use of the term 'hyperreal' — 'transferring it [human space] into a hyperreal of simulation' (Baudrillard, 1995, p. 82). It points to the ephemerality and insubstantiality of these additional social media users, whose association with actual, living individuals is precarious and will vary by profile and over dimensions of time and space (boyd, 2007). As such, the normative implications are distinct from those associated with overpopulation. It also signifies that the number of inau-

thentic users on social media may be much greater than has been reported to date.

Consequences of social media hyper-population and epistemic uncertainty

The consequences of social media hyper-population are not necessarily negative. As discussed above, inauthentic profiles can be beneficial to society, or simply immaterial to the functioning of the platform. Islamic prayer bots appear to be becoming a common part of Muslim culture (Öhman et al., 2019). The majority of social media profiles of people who have died accrete, but for the most part lie inactive and undisturbed. However, there are also problems associated with hyper-population and epistemic uncertainty about the nature and extent of inauthentic profiles.

The first issue concerns the value and business model of the companies themselves. As Elon Musk's lawyers wrote, 'Twitter's true mDAU count is a key component of the company's business, given that approximately 90% of its revenue comes from advertisements' (Ringler, 2022). The same holds true for all other major social media companies that rely on advertising for most of their income. Moreover, since much of the information economy on the web is based around attention, epistemic uncertainty about the authenticity of attention risks undermining the wider political economy of the internet. If advertisers come to believe that a large proportion of the attention and response they pay for is not real, then social media hyper-population may precipitate a 'subprime attention crisis' (Hwang, 2020). The accurate identification of inauthentic accounts is also a prerequisite to addressing the increasing problem of fraud online. False social media accounts have been found to have been used extensively to commit fraud, launch attacks and commit various financially motivated crimes online. The overall extent of this is significant and growing. In digital advertising, for example, the annual cost of fraud has been estimated at between $6.5 and $19 billion (Perrin, 2019). Other analyses have found that 28-40% of web traffic and clicks appear to be non-human, and used for the purpose of generating revenue from manufactured attention (Gordon et al., 2021).

In politics, social bots, sybils and compromised user accounts are used to manipulate issues and events. Since Facebook began reporting on 'inauthentic behavior' on its services, it has removed thousands of Facebook and Instagram accounts which it says were being used to 'manipulate public debate' in over 40 countries (Facebook Inc., 2018-2021). Former Facebook employee and whistleblower Sophie

Zhang claimed in a 2020 internal memo, and subsequently in an interview for this article, that this represents only a fraction of the inauthentic political activity (Silverman & Mac, 2020; Zhang, personal communication, 7 May, 2021). Other investigations and research appear to corroborate Zhang's claims. In 2019, for example, hired trolls 'worked round-the-clock to flood platforms such as Twitter and Facebook with seemingly organic messages of support' for a Senate candidate in the Philippines (Mahtani & Cabato, 2019). In the months leading up to the European Parliamentary elections in May 2019, false and suspect Facebook accounts were used to promote the far right AfD party (Davis et al., 2019). Networks of false profiles have been used to disseminate extremist propaganda in Germany (Baumgärtner & Höfner, 2020). Political consultancies have used false accounts to leak allegations about politicians and pose as local campaigners, such as the Israeli company the Archimedes Group, which was active across north and west Africa (Gleicher, 2019). Other public affairs companies have used such accounts to promote state propaganda (Davies, 2019). In each of these cases false accounts appear to have played a part in the distortion and manipulation of political issues or events, and the amalgamation of manufactured and genuine support, though there remains considerable uncertainty about the nature of the role they played.

Candidates and campaigns also hire or buy fake followers and engagement to give a false impression of popularity or representativeness. Estimates of the proportion of @realDonaldTrump followers who were false, for example, ranged from 3% to over 60% (McGill, 2016; Fishkin, 2018). In the lead up to the US 2016 Presidential election, a subsection of these bots linked to Russia retweeted @realDonaldTrump 469,527 times (Twitter Inc. Submission, 2019). In 2018 the Office of the New York State Attorney General estimated that up to 9.6 million comments lodged with the FCC regarding planned changes to net neutrality rules 'wrongfully used New Yorkers' identities without their consent' (James, 2018). Some of the identities used were those of people who had died (Singer-Vine & Collier, 2019). In Austria, a PR Agency called Mhoch3, employed people to post 80-100,000 comments per year under false online identities on behalf of institutional clients, including State-owned rail and bus companies (Apfl & Kleiner, 2014). Public support has, particularly in democratic societies, been seen as a justification for political action. Yet, if this public support is fabricated, or is perceived to be fabricated, then there is no such democratic justification.

The accumulation of additional social media accounts has also been associated with social problems. In some studies, users who are anonymous, or whose account is not linked to a single identifiable individual, have been found to be more

aggressive, and less likely to observe netiquette than profiles linked directly to a real author (Moore et al., 2012). Though other studies do not find such a link and point out that anonymity is a 'multifaceted construct' (Jaidka et al., 2022). Bots and sybils have been used for harassment and abuse (Ferrara, 2015), to amplify low-credibility content (Shao et al., 2018) and to post false reviews online (Competition and Markets Authority, 2021). Celebrities and commentators have been found to have artificially inflated their followers and engagements (Confessore et al., 2018). Even police have used false Facebook profiles to impersonate real people (Maass, 2019).

There is also a wider danger posed by social media hyper-population and epistemic uncertainty about the nature and extent of inauthentic profiles: the deterioration of public trust. Trust in the content of social media is already low, and will decline further if people believe that many of the users, and much of the engagement, is manufactured or counterfeit. Few people feel expert at assessing the authenticity of individuals online. A 2018 Pew Research Study found that only 7% of American adults were very confident they could recognize bots on social media, with 40% somewhat confident, and 53% not very or not at all confident (Stocking & Sumida, 2018). The public's lack of confidence regarding bots is coupled with their growing concern about bots online, and a perception that bots are a malignant influence (Stocking & Sumida, 2018). Left unaddressed, epistemic uncertainty could make social media look like a systemically untrustworthy space in which people are unsure who or what is real.

Conclusion

This article has shown that, though the leading social media companies recognise that a proportion of their users are inauthentic (in the sense of fake, spam or duplicate accounts), there is considerable uncertainty about the exact nature or number of these accounts. It has outlined three of the reasons for this uncertainty: the inconsistent definition of 'user', the low hurdles to account creation, and incomplete or irreconcilable calculations by social media companies. It has also sought to show how difficult it is for external observers to research the nature and extent of different user types, even on a platform as open and accessible as Twitter. For these reasons, there is limited existing knowledge about the number or type of inauthentic accounts on leading social media services. The figures that are available, and the research that has been done, suggest that the number of inauthentic accounts may be considerably different than the figures reported. These inauthentic accounts can, as this article also outlines, be the source of negative economic, po-

litical and social consequences. Though, due to epistemic uncertainty it is not possible to evaluate the degree to which they are the cause, or properly assess the impact they have.

The article therefore concludes that there is a need for detailed, consistent and assessable audits of the nature and extent of inauthentic accounts on major social media services. The companies claim, in their Securities' filings, that they already conduct such audits. Yet, as this article has sought to demonstrate, these audits are not sufficiently detailed, are inconsistent and are not properly assessable. Moreover, there are reasons to question the figures reported. There is therefore a compelling rationale for these audits to be conducted by an independent external body, and in such a way that they can be scrutinised by those outside the companies themselves. Similar independent audits have been proposed or conducted for digital intermediary algorithms, and for content moderation (Digital Regulation Cooperation Forum, 2022; Sarang, 2022).

Social media account audits are unlikely to result from current social media legislative initiatives in Europe and the US. These initiatives are directed chiefly at regulating content rather than users. Legislation passed or tabled, by the EU and the UK, to address online harms focuses on the sufficiency of intermediary processes for dealing with harmful content. Although both the EU's Digital Services Act (2020) and the UK's Online Safety Bill (2022) require risk assessments of platform services, there is no obligation to audit or enumerate social media accounts. The Digital Services Act (DSA) comes closest, in recognising that there may be particular risks associated with 'the creation of fake accounts, the use of bots, and other automated or partially automated behaviours', and requiring platforms to mitigate against such risks (Digital Services Act, 2020). Yet, the transparency requirements within the DSA, as with the transparency requirements contained within a raft of draft proposals put forward at the State and Federal level in the US (such as the Platform Accountability and Consumer Transparency Act [S.797], or the Platform Accountability and Transparency Act) have yet to specify transparency about the number or type of user accounts. For this reason, epistemic uncertainty about inauthentic accounts may not be resolved by current policy initiatives.

There is also a pressing need for further academic research on inauthentic accounts. This research will need to go beyond Twitter, where most research attention has been focused to date, to examine inauthentic accounts on Facebook, Instagram, Snapchat and other leading services. It will need to document who is creating inauthentic accounts, what accounts they are creating, how many accounts are they creating and for what purpose. Equally, there is a need for research that

analyses the accounts that are used to pursue coordinated inauthentic behaviour, in order to understand which accounts are being used for malignant or fraudulent purposes. This research will necessitate greater openness from the companies than we have seen to date, reiterating the need for further policy intervention.

References

Apfl, S., & Kleiner, S. (2014, November). Die Netzflüsterer [The Network Whisperers]. *Datum*. https://datum.at/die-netzfluesterer/

Bastos, M. T., & Mercea, D. (2019). The Brexit botnet and user-generated hyperpartisan news. *Social Science Computer Review*, *37*(1), 38–54. https://doi.org/10.1177/0894439317734157

Baudrillard, J. (1995). *Simulacra and simulation* (S. Glaser, Trans.). University of Michigan Press. https://doi.org/10.3998/mpub.9904

Baumgärtner, M., & Höfner, R. (2020, January 24). How to fake friends and iInfluence people. *Spiegel International*. https://www.spiegel.de/international/germany/facebook-how-to-fake-friends-and-influence-people-a-4605cea1-6b49-4c26-b5b7-278caef29752

Bay, S., & Fredheim, R. (2019). *How social media companies are failing to combat inauthentic behaviour online*(Social Media Manipulation Report 2019) [Report]. NATO Strategic Communications Centre of Excellence. https://www.stratcomcoe.org/how-social-media-companies-are-failing-combat-inauthentic-behaviour-online

Bessi, A., & Ferrara, E. (2016). Social bots distort the 2016 U.S. Presidential election online discussion. *First Monday*, *21*(11).

boyd, d. (2004). Friendster and publicly articulated social networking. *Proceedings of the Conference on Human Factors and Computing Systems (CHI 2004), April 24-29*, 1281–1282.

boyd, d. (2007). None of this is real: Identity and participation in Friendster. In J. Karaganis (Ed.), *Structures of participation in digital culture* (pp. 132–157). Social Science Research Council.

Competition and Markets Authority. (2021). *CMA intervention leads to further Facebook action on fake reviews* [Press release]. Competition and Markets Authority. https://www.gov.uk/government/news/cma-intervention-leads-to-further-facebook-action-on-fake-reviews

Confessore, N., Dance, G. J. X., Harris, R., & Hansen, M. (2018, January 27). The follower factory: Everyone wants to be popular online. *The New York Times*. https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html

Davies, C. (2019, November 1). Undercover reporter reveals life in a Polish troll farm. *The Guardian*. https://www.theguardian.com/world/2019/nov/01/undercover-reporter-reveals-life-in-a-polish-troll-farm

Davis, T., Livingston, S., & Hindman, M. (2019). *Suspicious election campaign activity on Facebook* (pp. 1–23) [Report]. School of Media & Public Affairs, The George Washington University. https://iddp.gwu.edu/suspicious-election-campaign-activity-facebook

Digital Regulation Cooperation Forum. (2022). *Auditing algorithms: The existing landscape, role of*

regulators and future outlook [Discussion Paper]. UK Government. https://www.gov.uk/government/publications/findings-from-the-drcf-algorithmic-processing-workstream-spring-2022/auditing-algorithms-the-existing-landscape-role-of-regulators-and-future-outlook

Digital Services Act, no. COM/2020/825 final, European Commission (2020). https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2020:825:FIN

Driscoll, K., & Walker, S. (2014). Big data, big questions. Working within a black box: Transparency in the collection and production of big Twitter data. *International Journal of Communication*, *8*(20), 1745–1764.

Echeverria, J., & Zhou, S. (2017). Discovery, retrieval, and analysis of the 'star wars' botnet in Twitter. *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, 1–8. https://doi.org/10.1145/3110025.3110074

Facebook Inc. (2012-2020). *Facebook annual reports* [Report]. Facebook Inc.

Facebook Inc. (2018-2021). *Facebook coordinated inauthentic behaviour reports* [Report]. Facebook Inc.

Facebook Inc. (2020). *Terms of service*. facebook.com/terms.php

Facebook Inc. (2021). *Community standards enforcement report* [Report]. Facebook Inc. https://transparency.fb.com/data/community-standards-enforcement/

Ferrara, E. (2015). Manipulation and abuse on social media. *ACM SIGWEB Newsletter*, *4*(9), 1–9. https://doi.org/10.1145/2749279.2749283

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, *59*(7), 96–104. https://doi.org/10.1145/2818717

Fink, A. (2010). *Conducting research literature reviews: From the internet to article* (3rd ed.). Sage.

Fishkin, R. (2018, October 9). We analyzed every Twitter account following Donald Trump: 61% are bots, spam, inactive, or propaganda. *SparkToro*. https://sparktoro.com/blog/we-analyzed-every-twitter-account-following-donald-trump-61-are-bots-spam-inactive-or-propaganda/

Fontanella-Khan, J., & Murphy, H. (2022, May 14). Why is Elon Musk really putting his Twitter deal 'on hold'? *Financial Times*. https://www.ft.com/content/6b84ddc6-63c8-4c4a-b1ff-582ac6a9f83c

Gehl, R. W. (2012). Real (software) abstractions: On the rise of Facebook and the fall of MySpace. *Social Text*, *30*(2(111)), 99–119. https://doi.org/10.1215/01642472-1541772

Gilani, Z., Farahbakhsh, R., Tyson, G., Wang, L., & Crowcroft, J. (2017). Of bots and humans (on Twitter). *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, 349–354. https://doi.org/10.1145/3110025.3110090

Gleicher, N. (2019). *Removing coordinated inauthentic behavior from Israel* [Press release]. Facebook Inc. https://about.fb.com/news/2019/05/removing-coordinated-inauthentic-behavior-from-israel/

Gleicher, N., Franklin, M., Agranovich, D., Nimmo, B., Belogolova, O., & Torrey, M. (2021). *Threat report: The state of influence operations 2017-2020* [Report]. Facebook Inc. https://about.fb.com/wp-content/uploads/2021/05/IO-Threat-Report-May-20-2021.pdf

Gordon, B. R., Jerath, K., Katona, Z., Narayanan, S., Shin, J., & Wilbur, K. C. (2021). Inefficiencies in digital advertising markets. *Journal of Marketing*, *85*(1), 7–25. https://doi.org/10.1177/0022242920913236

Gorwa, R., & Guilbeault, D. (2020). Unpacking the social media bot: A typology to guide research and policy. *Policy & Internet*, *12*(2), 225–248. https://doi.org/10.1002/poi3.184

Hofeditz, L., Bunker, D., Ehnis, C., Stieglitz, S., & Brachten, F. (2019). Meaningful use of social bots? Possible applications in crisis communication during disasters. *Proceedings of the 27th European Conference on Information Systems (ECIS)*. ECIS, Sweden.

Hwang, T. (2020). *Subprime attention crisis: Advertising and the time bomb at the heart of the internet*. Farrar, Straus & Giroux.

Instagram. (2021). *Terms of use*. Facebook Inc. https://www.facebook.com/help/instagram/termsofuse

Jaidka, K., Zhou, A., Lelkes, Y., Egelhofer, J., & Lecheler, S. (2022). Beyond anonymity: Network affordances, under deindividuation, improve social media discussion quality. *Journal of Computer-Mediated Communication*, *27*(1), zmab019. https://doi.org/10.1093/jcmc/zmab019

James, L. (2018). *Fake comments*. New York State Office of the Attorney General. https://ag.ny.gov/fakecomments

Keller, T. R., & Klinger, U. (2019). Social bots in election campaigns: Theoretical, empirical, and methodological implications. *Political Communication*, *36*(1), 171–189. https://doi.org/10.1080/10584609.2018.1526238

Kollanyi, B. (2016). Where do bots come from? An analysis of bot codes shared on GitHub. *International Journal of Communication*, *10*(2016), 4932–4951.

Lanier, J. (2008, August 8). *Big data commerce vs big data science. A conversation with Jaron Lanier* [Edge]. https://www.edge.org/conversation/jaron_lanier-big-data-commerce-vs-big-data-science

LinkedIn. (2016). *Annual report 2015* [Report]. LinkedIn Corporation. https://news.linkedin.com/2016/linkedin-announces-fourth-quarter-and-full-year-2015-results

Luceri, L., Deb, A., Giordano, S., & Ferrara, E. (2019). Evolution of bot and human behavior during elections. *First Monday*, *24*(9).

Maass, D. (2019). Four steps Facebook should take to counter police sock puppets. *Electronic Frontier Foundation Deeplinks*. https://www.eff.org/deeplinks/2019/04/facebook-must-take-these-four-steps-counter-police-sock-puppets

Mahtani, S., & Cabato, R. (2019, July 26). Why crafty internet trolls in the Philippines may be coming to a website near you. *The Washington Post*. https://www.washingtonpost.com/world/asia_pacific/why-crafty-internet-trolls-in-the-philippines-may-be-coming-to-a-website-near-you/2019/07/25/c5d42ee2-5c53-11e9-98d4-844088d135f2_story.html

Marwick, A. E., & boyd, danah. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, *13*(1), 114–133. https://doi.org/10.1177/1461444810365313

McGill, A. (2016, June 2). Have Twitter bots infiltrated the 2016 election? *The Atlantic*. https://www.theatlantic.com/politics/archive/2016/06/have-twitter-bots-infiltrated-the-2016-election/484964/

Moore, M. J., Nakano, T., Enomoto, A., & Suda, T. (2012). Anonymity and roles associated with aggressive posts in an online forum. *Computers in Human Behavior*, *28*(3), 861–867. https://doi.org/10.1016/j.chb.2011.12.005

Musk, E. [@elonmusk]. (2022a, May 13). *Twitter deal temporarily on hold pending details supporting calculation that spam/fake accounts do indeed represent less than 5% of users* [Tweet]. Twitter. https://twitter.com/elonmusk/status/1525049369552048129

Musk, E. [@elonmusk]. (2022b, May 14). *Any sensible random sampling process is fine. If many people independently get similar results for % of fake/spam/duplicate accounts, that will be telling. I picked 100 as the sample size number, because that is what Twitter uses to calculate <5% fake/spam/duplicate* [Tweet]. Twitter. https://twitter.com/elonmusk/status/1525304736538312707

Obar, J. A., & Wildman, S. S. (2015). Social media definition and the governance challenge: An introduction to the special issue. *Telecommunications Policy*, *39*(9), 745–750. https://doi.org/10.2139/ssrn.2647377

Oentaryo, R. J., Murdopo, A., Prasetyo, P. K., & Lim, E. P. (2016). On profiling bots in social media. In E. Spiro & Y.-Y. Ahn (Eds.), *Social Informatics. 8th International Conference, SocInfo 2016, Bellevue, WA, USA, November 11-14, 2016, Proceedings, Part I* (pp. 92–109).

Öhman, C., Gorwa, R., & Floridi, L. (2019). Prayer-bots and religious worship on Twitter: A call for a wider research agenda. *Minds and Machines*, *29*(2), 331–338. https://doi.org/10.1007/s11023-019-09498-3

Öhman, C. J., & Watson, D. (2019). Are the dead taking over Facebook? A big data approach to the future of death online. *Big Data & Society*, *6*(1). https://doi.org/10.1177/2053951719842540

Online Safety Bill, UK Government, as introduced 11 May 2022, 4 58 3 (2022). https://bills.parliament.uk/bills/3137/publications

Orabi, M., Mouheb, D., Al Aghbari, Z., & Kamel, I. (2020). Detection of bots in social media: A systematic review. *Information Processing & Management*, *57*(4), 102250. https://doi.org/10.1016/j.ipm.2020.102250

Perrin, N. (2019). *Digital ad fraud 2019. Mobile and video remain riskiest channels* [Report]. Insider Intelligence. https://www.emarketer.com/content/digital-ad-fraud-2019#page-report

Ringler, M. (2022, July 8). *Letter from Skadden, Arps, Slate, Meagher & Flom LLP to Twitter Inc*. https://www.sec.gov/Archives/edgar/data/1418091/000110465922078413/tm2220599d1_ex99-p.htm

Romanov, A., Semenov, A., Mazhelis, O., & Veijalainen, J. (2017). Detection of fake profiles in social media—Literature review. *Proceedings of the 13th International Conference on Web Information Systems and Technologies*, 363–369. https://doi.org/10.5220/0006362103630369

Roth, Y., & Harvey, D. (2018, June 26). How Twitter is fighting spam and malicious automation. *Twitter Blog*. https://blog.twitter.com/en_us/topics/company/2018/how-twitter-is-fighting-spam-and-malicious-automation.html

Roth, Y., & Pickles, N. (2020, May 18). Bot or not? The facts about platform manipulation on Twitter. *Twitter Blog*. https://blog.twitter.com/en_us/topics/company/2020/bot-or-not.html

Sarang, V. (2022). *Community standards enforcement report assessment results* [Report]. Meta. https://about.fb.com/news/2022/05/community-standards-enforcement-report-assessment-results/

Schultz, A. (2019). *How does Facebook measure fake accounts?* [Press release]. Facebook Newsroom. https://about.fb.com/news/2019/05/fake-accounts/

Shao, C., Ciampaglia, G. L., Varol, O., Yang, K.-C., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature Communications*, *9*(1), 4787. https://doi.org/10.1038/s4

1467-018-06930-7

Silverman, C., & Mac, R. (2020, December 10). Facebook gets paid. *Buzzfeed*. https://www.buzzfeednews.com/article/craigsilverman/facebook-ad-scams-revenue-china-tiktok-vietnam

Singer, D. and Project Therapy v Facebook, Class Action Complaint, Case 3:18-cv-04978 (US District Court for the Northern District of California 2018).

Singer-Vine, J., & Collier, K. (2019, October 3). Political operatives are faking voter outrage with millions of made-up comments to benefit the rich and powerful. *Buzzfeed*. https://www.buzzfeednews.com/article/jsvine/net-neutrality-fcc-fake-comments-impersonation

Snap Inc. (2021a). *Snap annual report 2020* [Report]. Snap Inc. https://s25.q4cdn.com/442043304/files/doc_presentations/presentation/2021/Snap-Inc.-2020-Annual-Report.pdf

Snap Inc. (2021b, November 15). *Terms of service*. https://snap.com/en-GB/terms

Stanford University. (2005, October 26). *James Breyer / Mark Zuckerberg Interview* [Zuckerberg Transcripts. 116]. https://epublications.marquette.edu/zuckerberg_files_transcripts/116

Stocking, G., & Sumida, N. (2018). *Social media bots draw public's attention and concern* [Study]. Pew Research Center. https://www.journalism.org/2018/10/15/social-media-bots-draw-publics-attention-and-concern/

Subrahmanian, V. S., Azaria, A., Durst, S., Kagan, V., Galstyan, A., Lerman, K., Zhu, L., Ferrara, E., Flammini, A., & Menczer, F. (2016). The DARPA Twitter bot challenge. *Computer*, *49*(6), 38–46. https://doi.org/10.1109/MC.2016.183

Tagarelli, A., & Interdonato, R. (2013). 'Who's out there?': Identifying and ranking lurkers in social networks. *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 215–222. https://doi.org/10.1145/2492517.2492542

Takacs, R., & McCulloh, I. (2019). Dormant bots in social media: Twitter and the 2018 U.S. senate election. *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 796–800. https://doi.org/10.1145/3341161.3343852

Thomas, K., Grier, C., & Paxson, V. (2012). Adapting social spam infrastructure for political censorship. *USENIX Workshop on Large-Scale Exploits and Emergent Threats*, 1–8. https://www.usenix.org/system/files/conference/leet12/leet12-final13_0.pdf

Trifiro, B. M., & Gerson, J. (2019). Social media usage patterns: Research note regarding the lack of universal validated measures for active and passive use. *Social Media + Society*, *5*(2). https://doi.org/10.1177/2056305119848743

Twitter Inc. (2013-2020). *Twitter annual reports* [Report]. Twitter Inc.

Twitter Inc Submission. (2019, January 19). *Update on results of retrospective review of Russian-related election activity. Submission to Senate Committee on the Judiciary*. https://www.judiciary.senate.gov/imo/media/doc/Edgett%20Appendix%20to%20Responses.pdf

Twitter Safety. (2020, June 12). Disclosing networks of state-linked information operations we've removed. *Twitter Blog*. https://blog.twitter.com/en_us/topics/company/2020/information-operations-june-2020

Varol, O., Ferrara, E., Davis, C. A., Menczer, F., & Flammini, A. (2017). *Online human-bot interactions: Detection, estimation, and characterization* (arXiv:1703.03107). arXiv. http://arxiv.org/abs/1703.03107

Yang, K., Varol, O., Davis, C. A., Ferrara, E., Flammini, A., & Menczer, F. (2019). Arming the public with artificial intelligence to counter social bots. *Human Behavior and Emerging Technologies*, *1*(1), 48–61. https://doi.org/10.1002/hbe2.115

Appendix

Interviews

Interviews were conducted for this article. Interviewees were chosen based on their responsibilities at one of the major social media companies, or on the basis of research they had conducted about social media profiles. All interviews were semi-structured and conducted by the author either in-person or via video conferencing (mode noted for each interview).

Interviewees

- Bay, Sebastian and Rolf Fredheim, NATO StratCom. Interviewed via MS Teams, 11 May 2020.
- Gorwa, Robert, Fellow at Institut für Internet und Gesellschaft in Berlin. Interviewed via MS Teams, 8 April 2020.
- Greenspan, Aaron, Think Computer Corporation. Interviewed via MS Teams, 11 May 2020.
- Öhman, Carl, Oxford Internet Institute. Interviewed in-person in Oxford, 12 March 2020.
- Roth, Yoel (Head of Site Integrity), Twitter. Interviewed via Google Meet, 17 June 2021.
- Stern, Peter (Director, Content Policy Stakeholder Engagement), Vishwanath Sarang, Bochra Gharbaoui, Sebastian Poehlmann, Meta/ Facebook. Interviewed via Zoom, 11 June 2021.
- Subrahmanian, V.S., Dartmouth College Distinguished Professor in Cybersecurity, Technology, and Society. Interviewed via MS Teams, 9 June 2020.
- Zhang, Sophie, former Facebook employee and whistleblower. Interviewed via MS Teams, 7 May 2021.

Early Registration Pages of Social Media Services (from the Internet Archive)

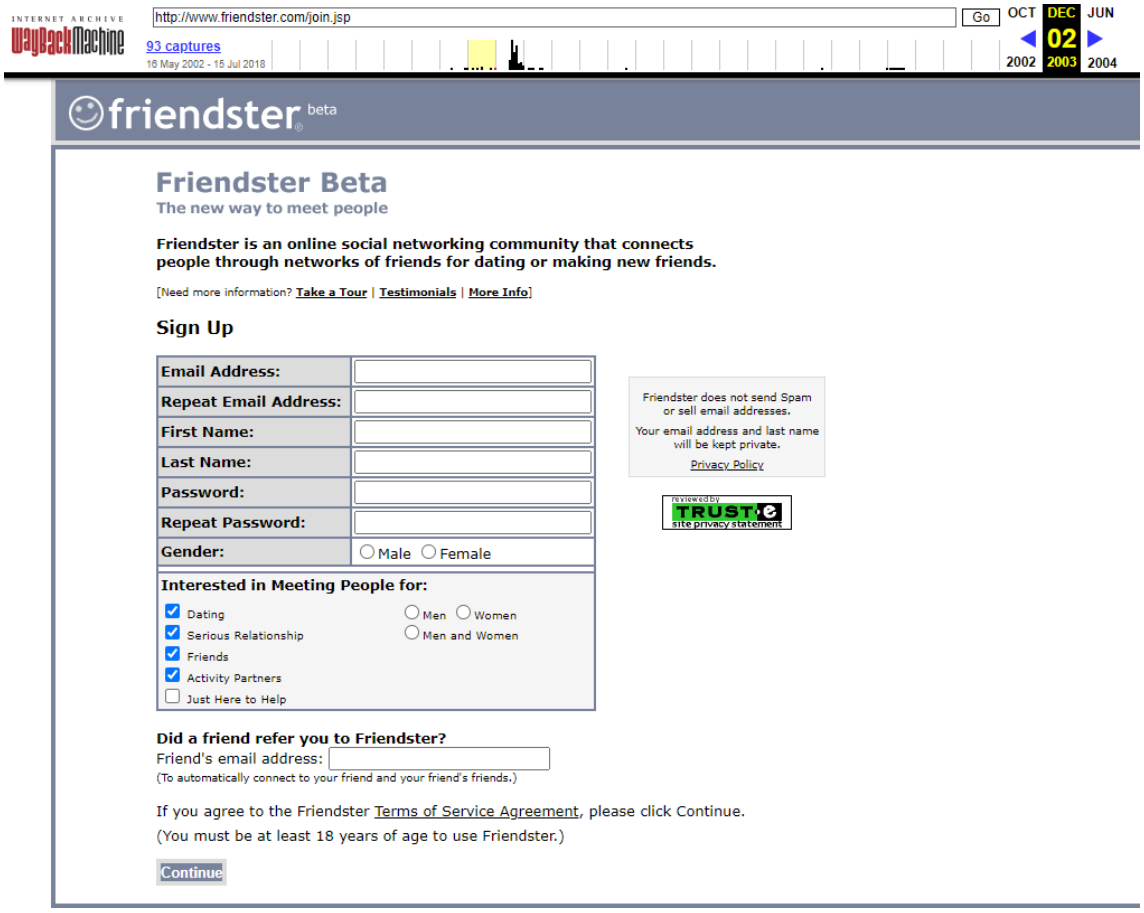

FIGURE 1: Friendster Registration Page (2 December 2003). Accessible via Wayback Machine.

FIGURE 2: MySpace Registration Page (19 September 2003). Accessible via Wayback Machine.

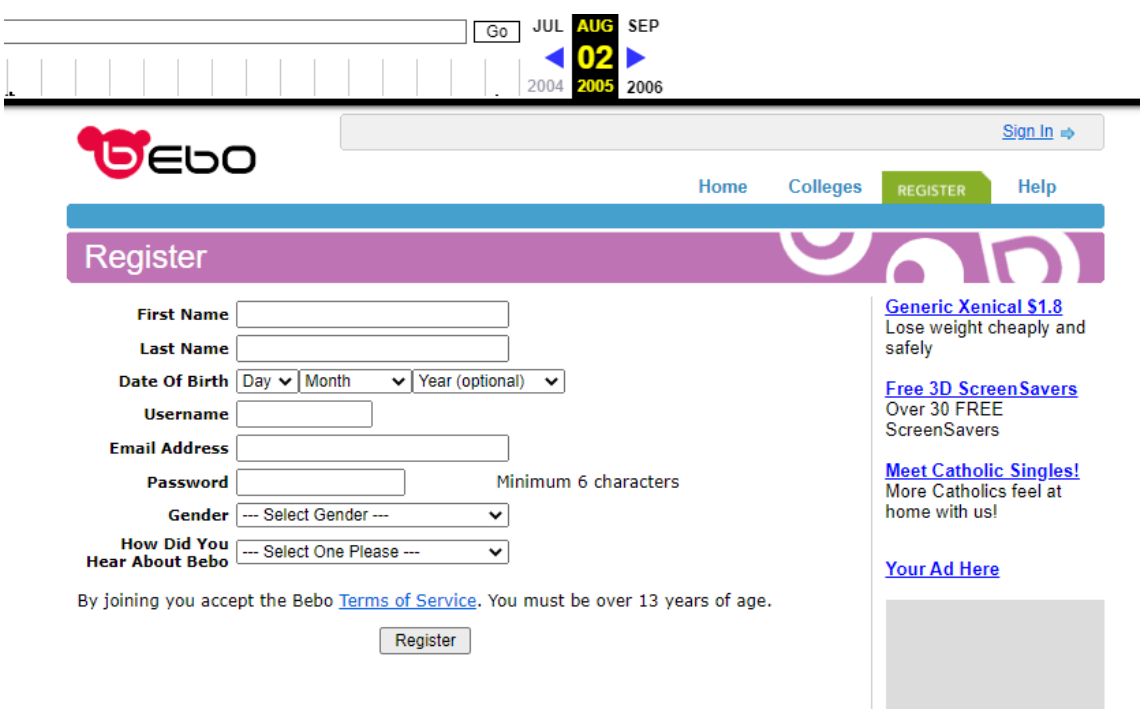

FIGURE 3: Bebo Registration Page (2 August 2005). Accessible via Wayback Machine.

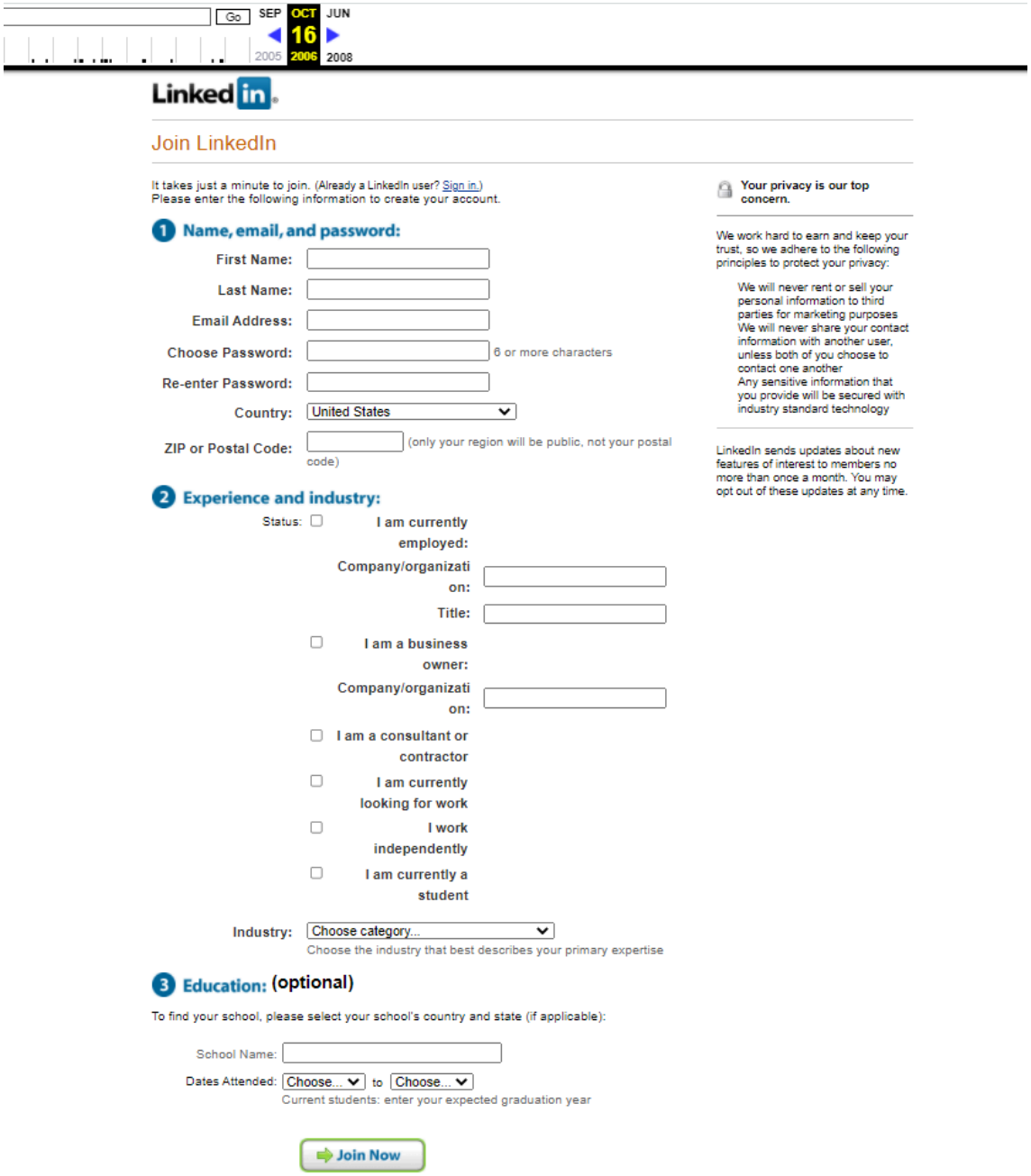

FIGURE 4: LinkedIn Registration Page (16 October 2006). Accessible via Wayback Machine.

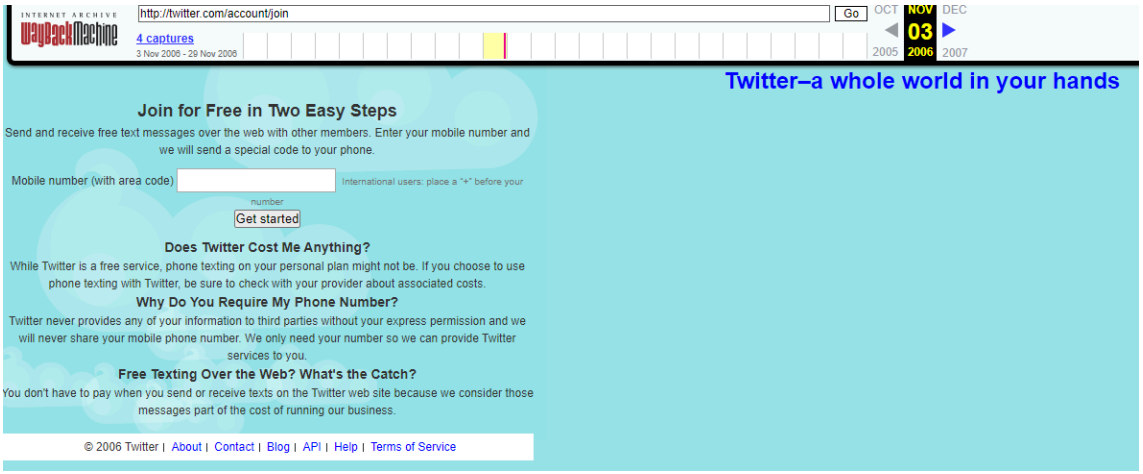

FIGURE 5: Twitter Registration Page (3 November 2006). Accessible via Wayback Machine.

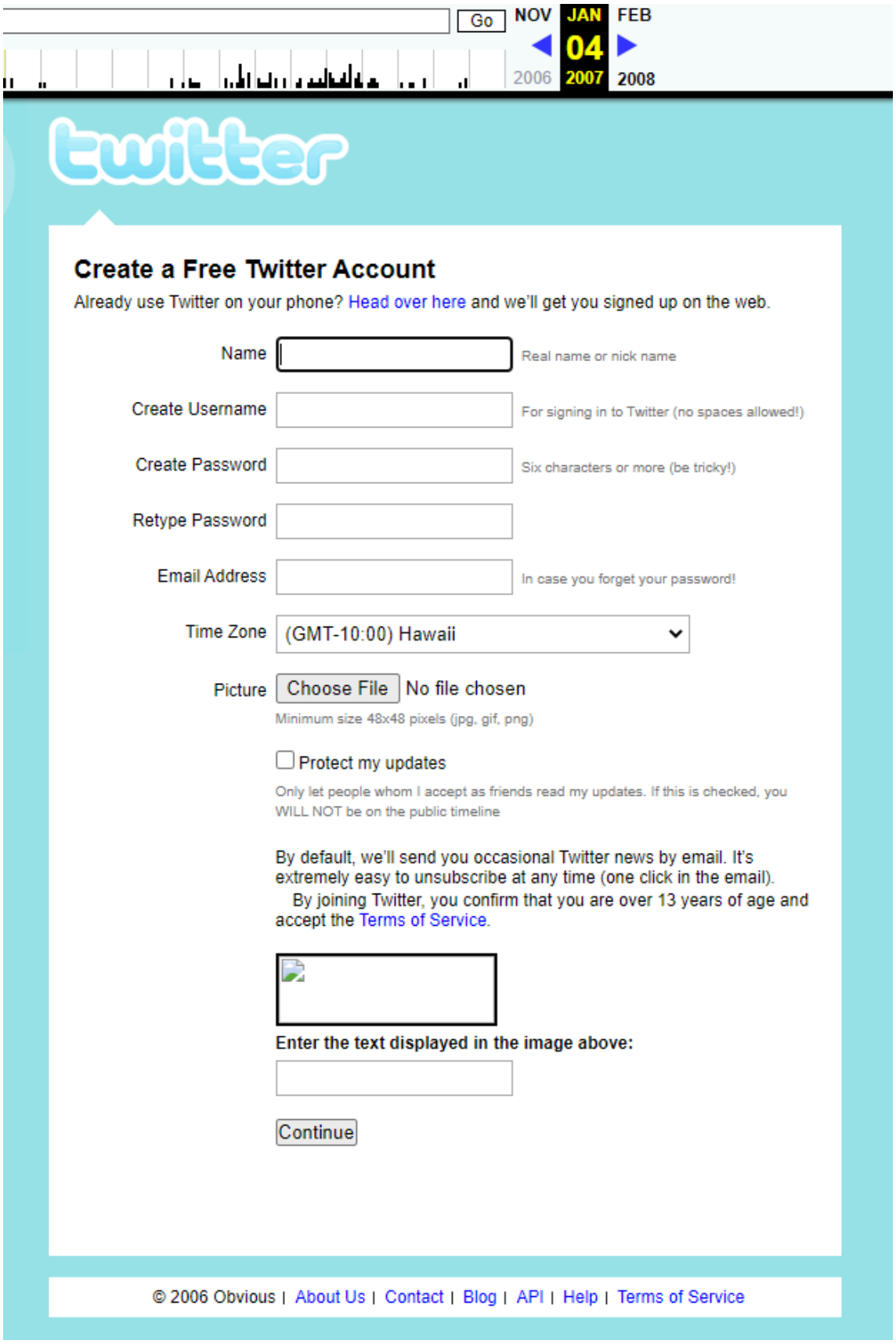

FIGURE 6: Twitter Registration Page (4 January 2007). Accessible via Wayback Machine.

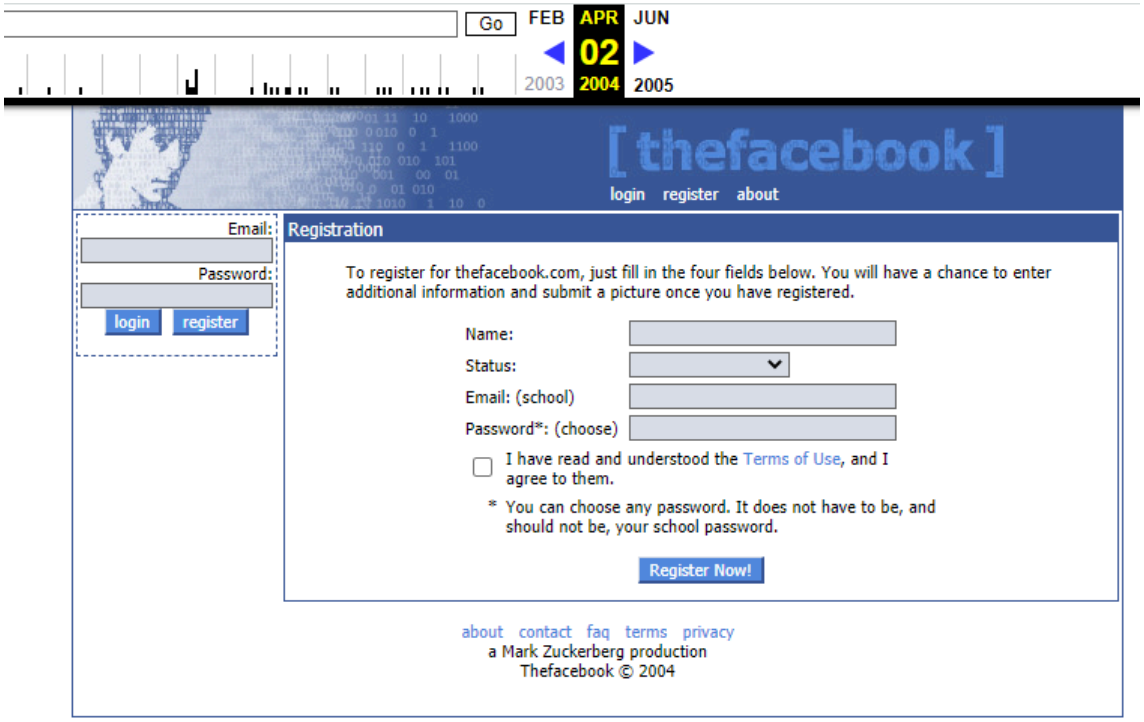

FIGURE 7: Facebook Registration Page (2 April 2004). Accessible via Wayback Machine.

FIGURE 8: Facebook Registration Page (13 August 2006). Accessible via Wayback Machine.

Published by

in cooperation with