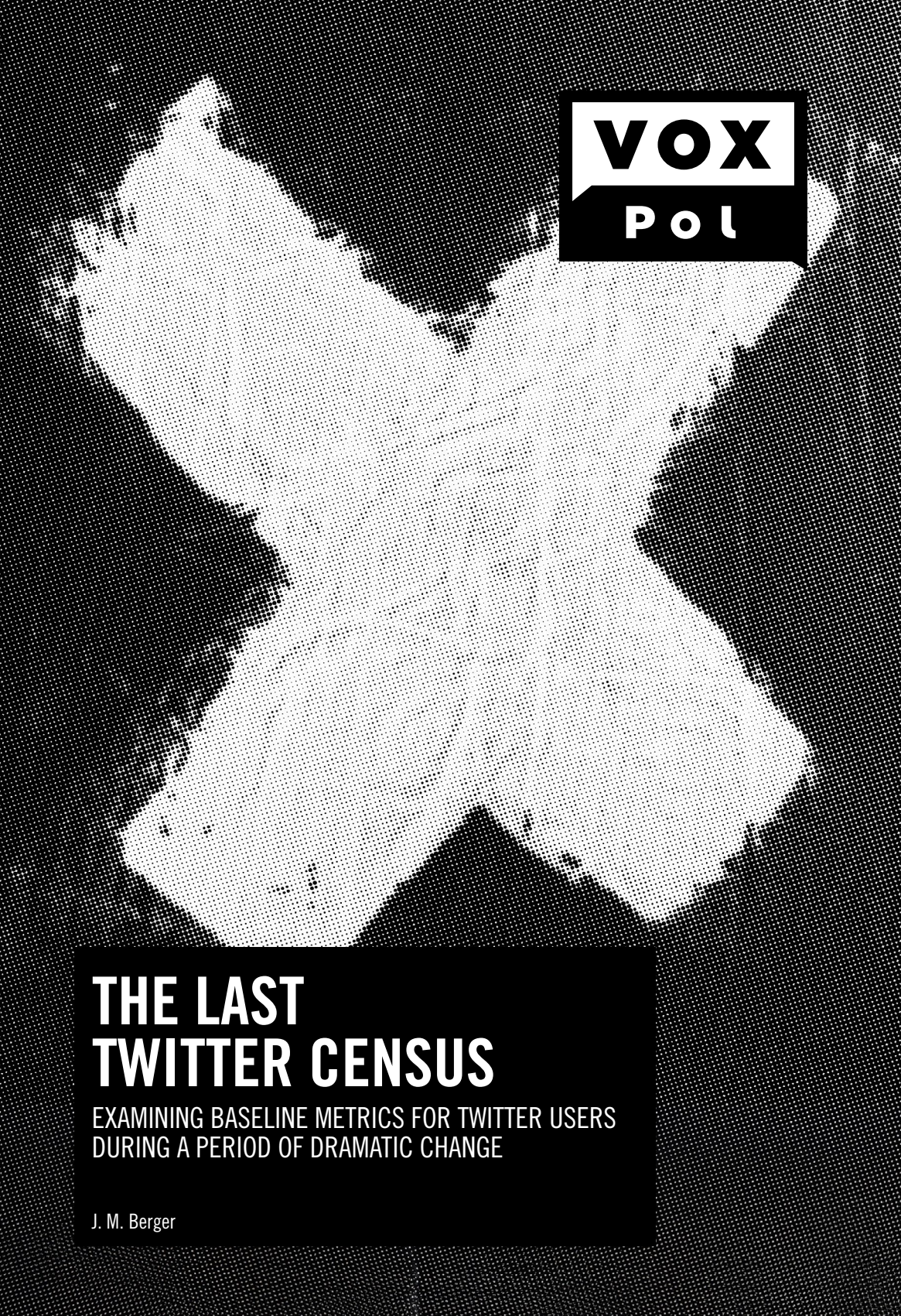


THE LAST TWITTER CENSUS

EXAMINING BASELINE METRICS FOR TWITTER USERS DURING A PERIOD OF DRAMATIC CHANGE

J. M. Berger

THE LAST TWITTER CENSUS

EXAMINING BASELINE METRICS FOR TWITTER USERS DURING A PERIOD OF DRAMATIC CHANGE

About the author

J.M. Berger is a writer and researcher focused on extremism as a Senior Research Fellow for the Center on Terrorism, Extremism, and Counterterrorism at the Middlebury Institute of International Studies. His research encompasses extremist and terrorist ideologies and propaganda, including social media and semantic analytical techniques. He is the author of four books, including Extremism (2018) and Optimal (2020). Berger is also a research fellow with VOX-Pol and a PhD candidate at Swansea University's School of Law, where he studies extremist ideologies. As a consultant for social media companies and government agencies, Berger conducts research and training on issues and policies related to homegrown terrorism, online extremism, advanced social media analysis, and countering violent extremism (CVE).

Acknowledgements

Special thanks to VOX-Pol's Dr. Alastair Reed, Prof. Stuart Macdonald and Louise Laing, and to Veronica Kelly, for shepherding this project to completion.

ISBN: 978-1-911669-63-0

Designed and typeset by Soapbox, www.designbysoapbox.com

Cover photo: Jen Theodore

TABLE OF CONTENTS

ABSTRACT

THIS REPORT COMPARES two large random samples of Twitter accounts that tweet in English: one taken just before Elon Musk acquired Twitter in October 2022, and one taken three months later, in January 2023. It also examines several related datasets collected during the period following the acquisition, a period in which, the study found, new accounts were created at a record-breaking pace. Some extremist and conspiracy networks created accounts faster than the baseline rate, probably because changes to Twitter's trust and safety policies had been announced. In the context of these policy announcements, the study examines some reinstated accounts, with mixed results. Despite the loosening of several content policies, accounts that automated the sending of tweets ('bots') saw activity drop sharply during the period of the study, with many bot accounts being suspended or deactivated, while others voluntarily curtailed their activity in light of the API changes announced. Deactivated accounts were dominated by sex-related content and apparent financial spam or scams, often coupled with automated tweeting.

EXECUTIVE SUMMARY

WHILE DISTINCT TRENDS emerged over the months covered by the study, the overall findings reflected a platform in the early stages of transformation. The landscape was shaped by policy announcements that sparked an influx of new users and a modest number of account deactivations.

Key findings include:

- The median account saw a decline in followers and friends (accounts followed), changes driven in part by a surge of new accounts created following Musk's acquisition of Twitter. Newer accounts generally tend to have fewer followers and friends.
- The median account tweeted less often – a change that was driven in part by the deactivation of low-hanging bot accounts, and was also consistent with reported declines in usage among some of the platform's most engaged users.
- Anti-LGBTQIA+ extremists and COVID-vaccine deniers were among the demographic groups overrepresented in the pool of new users, which skewed towards the American far right and support for Donald Trump. Accounts in both groups tweeted faster than the baseline account.
- While Twitter did reinstate thousands of previously suspended accounts, the reinstatements appeared to constitute only a small percentage of the total number suspended, based on a limited sample that comes with significant caveats. Not all reinstated users resumed tweeting.

1. INTRODUCTION

EVERY RESEARCH PROJECT is a journey. Some more than others.

This report examines baseline Twitter metrics from the period just before and just after Elon Musk's acquisition of the platform in October 2022.

My original goal was to drill down into a large sample of 'typical' Twitter accounts, with an eye to establishing a control group methodology and citable baseline metrics to inform comparative analyses. 'Typical' here refers to an attempt to abstract a representative portrait of how 'most' people use Twitter. Typical or baseline metrics refer, unless otherwise stipulated, to medians, which can serve as a starting point when comparing the set of all users to specific use cases, such as bot or spam accounts.

This replicable data approach was mainly intended to inform other research, such as the study of extremist movements, disinformation, content moderation practices and the evolution of the Twitter ecosystem over time. As a secondary focus, I wanted to collect data about content moderation practices in the broadest possible context, by examining account suspensions that took place within a large, randomised dataset.

The original objectives were overtaken by events when Elon Musk purchased Twitter on 27 October 2022, just days after the initial dataset was collected.

Musk immediately began to implement sweeping changes to the platform, including by firing a large swathe of the company's trust and safety team,[1] ending user verification in favour of a paid

1 Mike Masnick. "As Elon Fires More Trust & Safety Staff, Twitter's Moderation Efforts Fall Apart." Techdirt.com, 10 January 2023. www. techdirt.com/2023/01/10/as-elon-fires-more-trust-safety-staff-twitters-moderation-efforts-fall-apart/, retrieved 4 February 2023.

pseudo-verification product,[2] ending transparency initiatives[3] and reinstating a large number of accounts that had previously been suspended for various kinds of online harms, in a move that Musk characterised as a 'general amnesty'.[4]

By March 2023 Twitter insiders were warning that the company lacked the capacity to deal with the health of the online community, for example when it came to issues such as child sexual exploitation, harassment, extremism and state-sponsored disinformation.[5] Users were pushed to adopt an algorithmic feed,[6] and the algorithm was arbitrarily changed multiple times, resulting in less engagement for many users and excessive engagement for some.[7]

In short, typical Twitter usage in late 2022 and early 2023 was anything but typical.

The timeline and details of proposed changes were often unclear, creating a challenging environment for implementing the planned approach to enriching the original dataset. Just after the three-month

2 Lakshmi Varanasi. "Elon Musk said he's getting rid of the 'legacy' Twitter Blue badge, the longtime free verification for labeling famous and notable users." Insider, 3 February 2023. www.businessinsider.com/elon-musk-legacy-twitter-blue-will-soon-be-shut-down-2023-2, retrieved 4 February 2023.

3 Rob Pegoraro. "Twitter's transparency reporting has tanked under Elon Musk." Fast Company, 22 February 2023. www.fastcompany.com/90853794/twitters-transparency-reporting-has-tanked-under-elon- musk, retrieved 3 March 2023.

4 Rebecca Falconer. "Musk's Twitter 'amnesty' plan for suspended accounts alarms activists." Axios, 25 November 2023. www.axios.com/2022/11/25/musk-suspended-twitter-accounts-amnesty-plans-alarm-activists, retrieved 4 February 2023.

5 Marianna Spring. "Twitter insiders: We can't protect users from trolling under Musk." BBC.com, 6 March 2023. www.bbc.com/news/technology-64804007, retrieved 16 March 2023.

6 Ivan Mehta. "Twitter brings its 'For You' and 'Following' dual-timeline view to the web." TechCrunch, 14 January 2023. techcrunch.com/2023/01/13/twitter-brings-its-for-you-and-following-dual-timeline-view-to-the-web, retrieved 16 March 2023.

7 Zoë Schiffer and Casey Newton. "Yes, Elon Musk created a special system for showing you all his tweets first." Platformer, 14 February 2023. www. platformer. news/p/yes-elon-musk-created-a-special- system, retrieved 16 March 2023.

anniversary of the acquisition, Twitter announced it would end free access to the public API, almost immediately killing off a number of long-standing third-party Twitter clients.

The API change also included the termination of an academic API access programme, creating major new obstacles to future research and, in particular, threatening to disrupt data collection via the API.[8] This report did not use the academic API, employing instead the public API and complying with its (considerable) rate limit restrictions. In other words, collection was very slow, and every effort to collect data at scale risked being terminated midstream.

Changes to the technical infrastructure and business model also raised serious questions about the viability of the future research this report was meant to inform.[9] At the end of March 2023 Twitter announced that it would charge $42,000 a month for very limited access to the API – prohibitively expensive for most academic research. The company also faces other problems that threaten its continued existence, including steep drops in advertising, the underperformance of new revenue sources, and potentially devastating clashes with regulators over its policing of child sexual abuse materials (CSAM), disinformation and extremism, especially in Europe.[10]

8 Ryan Browne. "Twitter will start charging developers for API access as Elon Musk seeks to drive revenue." CNBC.com, www.cnbc.com/2023/02/02/twitter-to-start-charging-developers-for-api-access.html, retrieved 4 February 2023.

9 Adam Conner. "The Dangers of a Twitter Bankruptcy or Acquisition."AmericanProgress.org, 25 January 2023. www. americanprogress. org/article/the-dangers-of-a- twitter-bankruptcy-or- acquisition, retrieved 16 March 2023.

10 Chris Stokel-Walker. "Twitter's $42,000-per-Month API Prices Out Nearly Everyone." Wired, 18 March 2023. www.wired.com/story/twitter-data-api-prices-out-nearly-everyone; Ryan Mac and Tiffany Hsu. "Twitter's U.S. Ad Sales Plunge 59% as Woes Continue." New York Times, 5 June 2023. www.nytimes.com/2023/06/05/technology/twitter-ad-sales-musk.html; Sam Schechner. "Twitter to Face Stress Test This Month, Top EU Tech Regulator Says." Wall Street Journal, 1 June 2023. www.wsj.com/ articles/ twitter-to-face-stress-test-this-month-top-eu-tech-regulator-says-62a7b64a.

All of these challenges could be boiled down to three existential questions:

- Would Twitter continue to exist?
- Would future researchers be able to access the API?
- Would the original set of user metrics remain meaningful after a barrage of drastic policy and technical revisions?

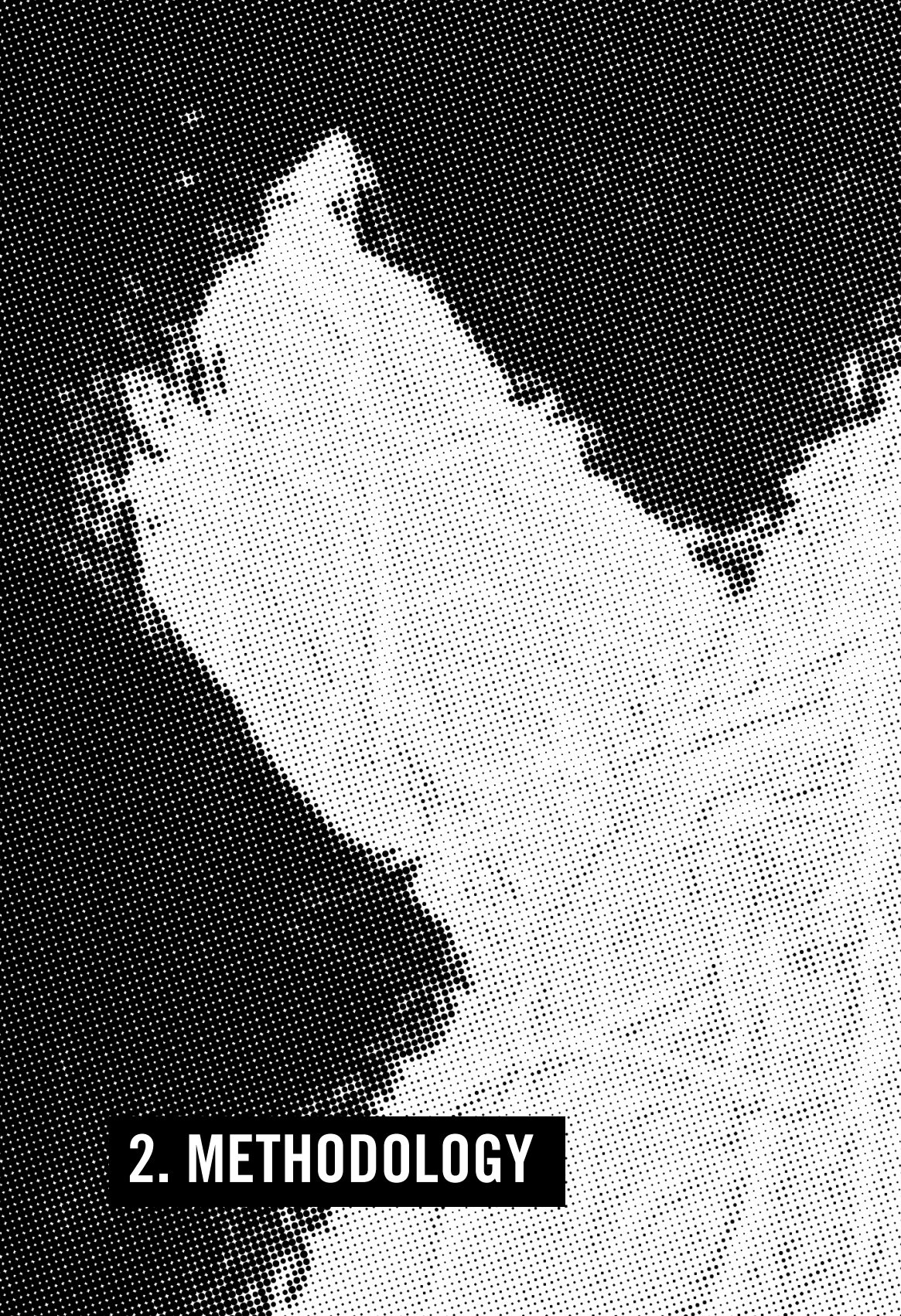
In response to the changing conditions, I updated my research questions to optimise my remaining time with free access to the API. This report will now explore the following questions:

- What were the baseline metrics for a typical account based on a large dataset of accounts that tweeted in English before and after Elon Musk's acquisition of Twitter?
- How did baseline metrics evolve in the months following the acquisition?
- How did Twitter policy announcements/changes affect baseline metrics?
- How did baseline accounts compare with accounts that took part in harmful behaviour such as anti-LGBTQIA+ extremism and the dissemination of anti-vaccination conspiracy theories?
- How many users were suspended or deactivated during the period of the study, and how did these users compare with the baseline?

This exploration will examine the following datasets:

- A large random sample of users who tweeted in English in October 2022 (hereafter the **October dataset**).
    - A subset of this dataset comprising accounts that were deactivated or suspended between October 2022 and February 2023 (the **deactivated dataset**)
    - Users from the October dataset were resampled for updated tweet data in December 2022 (the **December tweet dataset**).
    - Users from the October dataset were resampled to track user evolution in the February to March 2023 time frame (the **March resample dataset**).
- A comparable, large random sample of users who tweeted in English in January 2023 (the **January dataset**).
- A medium-sized sample of accounts that tweeted hashtags relating to disinformation about COVID vaccines in February 2023 (the **anti-vaccination dataset**). Users in this dataset were identified by their use of anti-vaccination hashtags.
- A medium-sized sample of accounts that tweeted anti-LGBTQIA+ content in February 2023 (the **groomer-slur dataset**). Users in this dataset were identified through their use of the anti-LGBTQIA+ slur 'groomer'.
- Three smaller datasets of accounts that were previously suspended and subsequently reinstated owing to recent changes to Twitter policies (the **reinstated dataset**). These datasets were generated by resampling lists of suspended accounts that had previously been identified by the author.

Note: After this report was written, Twitter changed its name to X. For the sake of simplicity, the platform is referred to herein as Twitter.



2. METHODOLOGY

WHAT IS A typical Twitter user, and how can we identify one?

The question is not as simple as it may seem. People use Twitter in many different ways. Some tweet regularly about a variety of subjects, others tweet mostly about one particular topic. Some people never tweet, using the platform only to read news. Some tweet about niche interests (such as sports), with high volumes during key events (such as a championship game) and low volumes at all other times. Some accounts tweet professionally or on behalf of institutions, while others are vehicles for personal expression. Some tweeters are not people at all, but automated bots (created by people). All these and more could be considered 'typical' use cases, but no one group is representative of the whole.

In an attempt to create a broadly representative sample, I set two very basic parameters for inclusion. To be part of this study, an account must tweet or retweet, and it must tweet or retweet in English at least some of the time.

- I chose to examine *accounts that tweet*, as opposed to accounts that never tweet, since accounts that tweet can be compared meaningfully with other accounts that tweet, such as users who employ a particular term or hashtag.
- I chose to examine *tweets in English*, my native language, because the methodology I chose for randomising the sample requires at least one member of a research team to be very fluent in the language being studied, having a firm grasp of grammar, word usage, slang, Internet-specific slang and other vernacular.

To identify a random sample of accounts that tweet in English, I used the Twitter search API to identify tweets containing some of the most common and generic English words.

According to the Oxford English Dictionary and corpus,[11] the most frequently used English words are:

1. The
2. Of
3. Be
4. And
5. To
6. In
7. A
8. That
9. Have
10. For

I removed some words from consideration (the, be, to, that, for) because they are often abbreviated or replaced with slang ('da' in the place of 'the,' 'b' in the place of 'be,' '2' in the place of 'to,' and so forth). Including these words might skew the demographics of responsive tweets, for instance according to a user's age or the Internet subculture to which they belong. I also removed 'a' after a sample search captured a very large number of non-English tweets. The final list of search terms included:

- Of
- And
- In
- Have

11 OED Online, Oxford University Press, December 2022, www.oed.com, retrieved 4 February 2023.

The targeted size of the dataset was >100,000 accounts, a robust sample that could be collected and analysed within the time allotted for the project.

The initial dataset was identified using a series of Twitter searches between 10 October 2022 and 26 October 2022. Following Twitter's API restrictions at the time, each search returned about 3,000 tweets, which were collected using the web service Twitonomy. Search times were staggered in order to sample at different times of day and on different days of the week. Searches were also spaced out in an effort to reduce the impact on the sample of particular events such as sports broadcasts, natural disasters or mass shootings. Searches captured both original tweets and retweets. The search results were collated, and duplicates removed. A total of 110,407 accounts were identified in the first round.

A custom Python application was then used to collect additional information about the accounts collected, including each user's full profile data and their 200 most recent tweets.[12] This collection process took some weeks, to comply with the API's rate limits, and was carried out in batches between 11 October and 26 October.

User profiles in the original dataset were resampled in February 2023, to determine how many accounts had been suspended or self-deleted (the deactivated dataset) and how user metrics had otherwise evolved. Tweets from the users in the original dataset were resampled from late February to early March, to capture how tweeting behaviour had evolved (the March resample dataset).

A second large dataset, randomised using the same method as for the October one, was identified in searches that took place between 5 January 2023 and 16 January 2023. Search collection times were again staggered, then tweets and profile information were collected in batches between 9 January and 17 January. A total of 108,723 accounts were identified.

Pertinent data collected for each type of account included:

- Number of followers
- Number of friends (accounts that a user follows)

12 Twitonomy.com. The Python app was developed by J.M. Berger, Dan Sturtevant and Jonathon Morgan.

- Number of tweets
- Tweets per day
    - Based on the last 200 tweets
    - Based on the lifetime of the account
- Twitter profile data (URL, location, hashtag, custom avatar)
- Date created (account age based on last recorded tweet)

Two different methods were used to calculate tweets-per-day, in order to gain a better understanding of current conditions and user evolution. Tweets per day (last-200 method) is generally a snapshot of how rapidly an account was tweeting at the time the sample was taken. Tweets per day (lifetime method) gives a picture of how the account tweeted over a longer period of time. As discussed below, both metrics bring value to the analysis, allowing for inferences about changes in behaviour.

The collected data was analysed, and the results are discussed in the following sections.

Unless otherwise specified, the baseline metrics discussed herein are median values based on the entire dataset. Mean averages tended to skew misleadingly high, owing to the presence of significant outliers in all of the datasets. For instance, eight accounts in the October dataset had more than 10 million followers, with the highest (@nytimes) having 54,601,797 followers and the second- highest (@fcbarcelona) having 44,595,277.

There was some noise in the dataset, particularly around tweets per day. Some accounts were incorrectly assigned a zero value for tweets per day (last-200 method), partly owing to three factors:

1. The Twitter API returned an increasing number of error responses during the period of the study,[13] adding a small amount of noise to the tweets per day (last-200 method) and showing some accounts as deactivated when they were not. Except where noted, API errors were immaterial.

2. A small number of accounts deleted their tweets between the time the account was identified and the time tweets were collected. Deleted tweets were not readily distinguishable from API errors. This issue affected only the last-200 method, as the lifetime tweets value returned by the API includes deleted tweets.

3. Some accounts changed their settings from public to private, which meant that they were initially detected tweeting publicly but they subsequently locked their accounts so that only followers could see their tweets. As a percentage of all accounts, this metric increased notably during the study period.

The increase in the number of private accounts was the primary driver of zero-tweets results, soaring from 0.3 per cent of the dataset to 1.4 per cent in January and 3 per cent by March. This large shift may be connected to increased abuse, harassment or toxic content, but there was also a notable period in late January/early February during which a significant number of high-profile users claimed they could get more engagement by locking their accounts.[14]

13 Ryan Mac, Mike Isaac and Kate Conger, 'Sometimes Things Break': Twitter Outages Are on the Rise." New York Times, 28 February 2023. www. nytimes.com/2023/02/28/technology/twitter-outages-elon-musk.html, retrieved 16 March 2023; Josephine Watson. "Twitter's daily limit outage proved to me the platform is dying." TechRadar, 9 February 2023. www. techradar. com/features/twitters-daily-limit-outage-proved-to-me-the-platform-is-dying, retrieved 16 March 2023.

14 Saksha Menezes. "Elon Musk Locks His Twitter Account to Test Engagement Impact." Bloomberg, 1 February 2023. www.bloomberg.com/news/articles/2023-02-01/musk-locks-his-twitter-account-to-test- engagement-impact, 16 March 2023.

Taking all of these factors into consideration, tweets-per-day scores of zero were excluded from the median calculation for the last-200 tweets method. The lifetime tweets-per-day metric was unaffected, as the API returns the total number of tweets even if an account is locked. In addition, accounts that were less than one day old were excluded from the lifetime metric, as fractional day calculations produced problematic results (e.g., someone who opened an account and tweeted 10 times in the first hour was credited with 240 tweets per day).

3. BASELINE USER METRICS

IN THIS SECTION I will examine the profile of a typical Twitter account that tweets in English, as captured in two rounds of data collection: one before and one after the acquisition of Twitter by Elon Musk. As noted above, the first data set was captured in October 2022 and the second in January 2023.

Table 1: Baseline user metrics, October 2022 versus January 2023

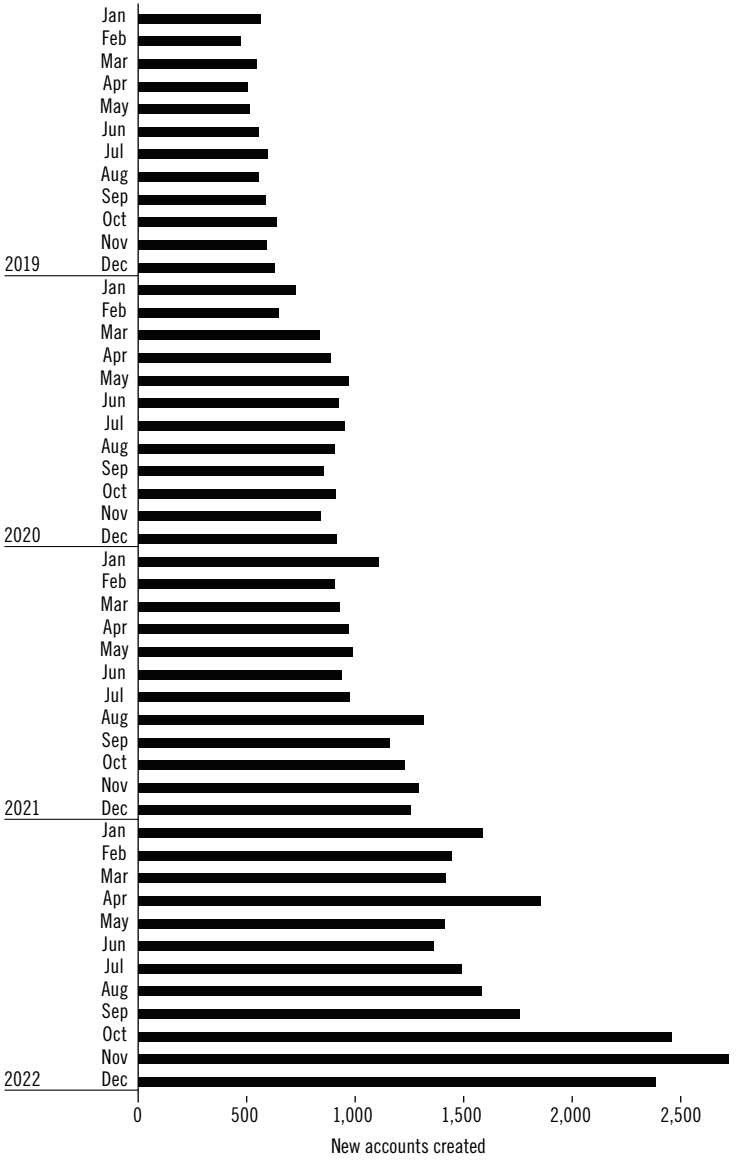
METRIC	OCTOBER 2022	JANUARY 2023	% CHANGE
# of accounts (n)	110,407	108,723	-1.53%
Median age (days)	1,531.31 min: 0 max: 5,830	1,583.27 min: 0 max: 6,029	3.39%
Median friends	474 min: 0 max: 1,503,514	441 min: 0 max: 911,148	-6.96%
Median followers	302 min: 0 max: 54,601,797	289 min: 0 max: 25,703,999	-4.30%
Median tweets/day, last 200 tweets[15]	22.22 min: 0.001 max: 200	18.18 min: 0.001 max: 200	-18.18%
Median tweets/day, lifetime	9.38 min: 0.001 max: 3,461.34	8.87 min: 0.004 max: 18,381.66	-5.50%
Median # of tweets, lifetime	10,724 min: 1 max: 15,307,992	10,404 min: 1 max: 52,005,200	-2.98%

The baseline user profile changed considerably during the three months between collections (Table 1). In January 2023, the median account had 441 friends and 289 followers, down 6.96 per cent and 4.30 per cent respectively from October. While a number of factors probably contributed to this decline, including users leaving the platform owing to increased harassment or in protest at the new

15 The last-200 method produced a maximum tweets-per-day score of 200. Users who tweeted more than 200 times per day were scored as 200.

management, most of the change can be attributed to a surge in new accounts (detailed below) – this exerted downward pressure, as newer accounts require time to build robust networks. Excluding accounts younger than 90 days completely erased the declines.

Figure 1: New account creation per month since 2019



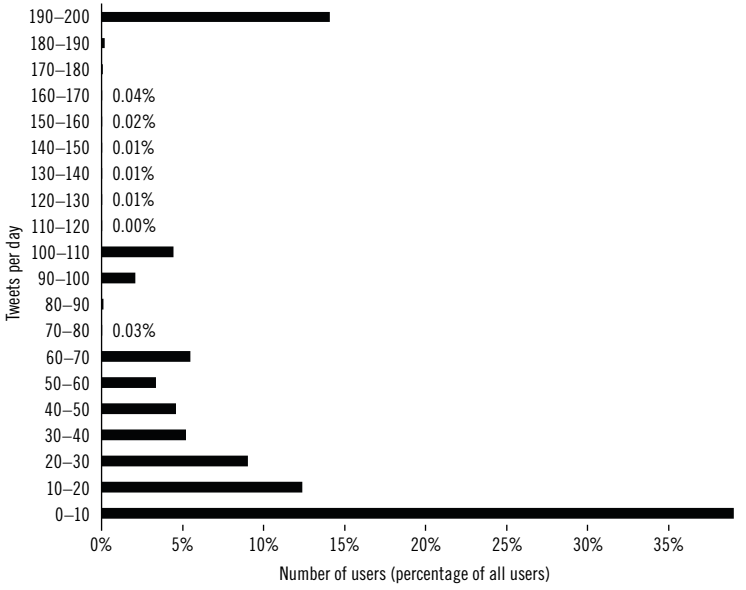
Source: January 2023 dataset

Account age increased by 3.39 per cent. This metric was calculated by measuring the number of days between the user's most recently recorded tweet (when available) and the account creation date. Because accounts were first identified in a search, and tweets were collected later, this metric undercounted very new (< 3 weeks) accounts. Accounts that changed their settings to 'private' in between identification and collection (332 in October, 1,490 in January) were excluded from the median, as the date of their last tweet could not be determined. The increase in median account age primarily reflected the passage of time across the long tail of the dataset, which was sufficient to offset a surge in new accounts.

Account creation spiked to a record high on 28 October 2022, the day after Musk purchased the platform – far higher than on any other single day in the entire January dataset. The spike followed three months of noteworthy but gradual growth. The second-best day for account creation was 26 April 2022, when Musk first formally announced his intention to buy Twitter.

The pace of new account creation remained elevated throughout November 2022, then began to slow in mid-December, the last full month for which data was collected. Even with the slowdown, October, November and December 2022 recorded the three highest tallies of account creations for any month on record (Figure 1). The median user's lifetime number of tweets sent dropped by nearly 3 per cent, driven primarily by this increase in new accounts.

Figure 2: Tweets per day, by range, January dataset



The median tweets-per-day metric in January was 18.18 (last-200 method), down 18.18 per cent from 22.22 in October. Lifetime tweets per day also declined, dropping to 8.87 from 9.38 in October, a decline of 5.50 per cent.

While more representative than the mean, median tweets per day painted a somewhat misleading portrait of typical use (Figure 2). In January, a plurality of accounts (about 39 per cent) tweeted fewer than 10 times per day, making this lower level of usage more 'typical' than the median. A smaller but still significant bloc of users (about 14 per cent) tweeted over 200 times per day. Almost no one tweeted more than 120 tweets per day but less than 200.

The number of prolific tweeters declined sharply between October and January. In October, 15.45 per cent of the dataset tweeted over 200 times per day (last-200 method), and just one tenth of 1 per cent (131 accounts) tweeted a thousand times or more per day (lifetime method). As previously noted, the last-200 method caps tweets per day at 200.

In January, only 10.78 per cent of the dataset tweeted 200 times or more per day (last-200), compared with 15.45 per cent in October. Accounts tweeting a thousand times or more per day (lifetime) remained level, at one tenth of 1 per cent (110 accounts).

Prolific tweeters showed signs of automated activity, and the decline in tweets-per-day probably reflects measures to rein in bot activity on the platform, although many bots continued to function as of this writing. This question will be examined in more detail under the heading 'Automated tweeting and spam'. The decline in tweeting would also be consistent with published reports citing internal Twitter documents that pointed to declining participation by some of the platform's most engaged users.[16]

I analysed the language each user included in the 'bio' field of their Twitter profile, typically a few short phrases or sentences of autobiographical exposition. Almost 83 per cent of the January dataset included at least one character in the bio field, down slightly from about 84 per cent in October. Using Sketch Engine, I produced a list of keywords and terms[17] which are words and short phrases that appear more often in the selected corpus than in the standard English corpus. All words in the bios were converted to lowercase, as SketchEngine sometimes makes distinctions based on capitalisation, and many Twitter users capitalise words in unexpected ways.

The 20 most-used profile keywords in the January dataset included 'fan', 'lover', 'tweet', 'writer', 'proud', 'dm' (short for direct message), 'enthusiast', 'mom', 'dad' and 'Twitter'. As a cautionary note, even some of the most-used keywords still appeared infrequently on an absolute basis, especially after the top ten. For instance, the text strings 'crypto' and 'nsfw' (not safe for work)

16 Sheila Dang. "Exclusive: Twitter is losing its most active users, internal documents show." Reuters, 25 October 2022. www.reuters.com/technology/exclusive-where-did-tweeters-go-twitter-is-losing-its-most-active-users-internal-2022-10-25.

17 Keywords are terms that appear more often in a corpus than in regular use. The reference corpus for this analysis was English Web 2020 (enTenTen20). Analysis was performed in Sketch Engine. For more information, see www. sketchengine.eu/my_keywords/keyword/#:~:text=Keywords%20are%20words%20(single%2Dtoken,reference%20corpus)%20or%20its%20subcorpus, retrieved 25 February 2023.

appeared frequently relative to other keywords, but rarely in absolute terms, featuring in 985 and 688 profiles respectively out of the 90,000 or so profiles that included English words.

Top multiword phrases included ‘fan account’, ‘official twitter’, ‘video game’ and ‘content creator’. A number of accounts included the phrase ‘minors dni’ (short for “do not interact”), generally reflecting the presence of sexual content. Other common multiword strings such as ‘are not endorsements’ and ‘are my own’ were consistent with widely used disclaimers included by professionals who tweet.

The text string ‘*suspen*’ appeared in 223 January profiles, down from 299 in October. The string almost always occurred in reference to accounts that claimed they had previously been suspended, often multiple times.

Accounts created in the weeks following Musk’s acquisition of Twitter showed a shift in profile language. The top 20 list was still dominated by generic terms such as ‘Twitter’, ‘account’, ‘fan’, ‘stan’ and ‘lover’. But more distinctive keywords emerged, including ‘conservative’, ‘patriot’, ‘MAGA’, ‘anti’ and ‘crypto’. The word ‘pronoun’ also rose into the top 20, used in both positive and negative contexts relating to gender identification.

By March 2023, 13.7 per cent of users created after the acquisition had already been deactivated, whether through suspension or self-deletion – a much higher rate than the 6.82 per cent seen in the overall October dataset. Account deactivations are detailed in Section 4.

3.1 USER EVOLUTION

Table 2: Results from resampling the original dataset, October 2022 versus February/March 2023

METRIC	OCTOBER 2022	MARCH 2023	% CHANGE
# of accounts (n)	110,407	102,422	-7.23%
Median friends	474	544	14.77%
Median followers	302	371	22.85%
Median tweets/day, lifetime	9.38	9.85	4.93%
Median # of tweets, lifetime	10,724	14,653	36.64%

In addition to collecting two comparable random samples, I wanted to examine how existing users evolved over time. From late February into early March 2023, as time was running out on the free API, I resampled the original October dataset.

The collection job was completed on 10 March 2023, but a large number of errors were returned by the API, repeatedly interrupting collection and ultimately corrupting the database of tweets. These errors were probably connected to an overall decline in site reliability reported elsewhere.[18] Because of this, I was unable to calculate accurate tweets-per-day scores using the last-200 tweets method for the resampled data, nor was I able to calculate the number of retweets or replies. I was, however, able to recover the last tweet date for each account, which made it possible to calculate tweets per day using the lifetime method. Tweets per day (lifetime) rose by almost 5 per cent. Friends and followers increased by 14.77 per cent and 22.85 per cent respectively. While directly comparable data for earlier

18 Clare Duffy. "Twitter hit with one of the biggest outages since Elon Musk took over." CNN, 6 March 2023. www.cnn.com/2023/03/06/tech/twitter-website-down/ index.html, retrieved 17 March 2023.

periods was not available, much of the growth would be consistent with the expected natural consequence of continued, active participation on the platform.

Locked accounts (with tweets viewable only by followers) soared from 0.3 per cent of the dataset to 3.0 per cent between October and March. As noted above, users may have been setting their tweets to private in order to increase their engagement. Given the size of the shift, however, it is also worth noting that various media reports pointed to increased harassment on the platform after October, which could be a factor.[19]

Other metrics pointed to newer users becoming more comfortable with the platform and its features. For instance, the number of accounts whose profile contained bio information, a custom photo avatar and a website link, increased modestly.

19 Stephen Neukam. "LGBTQ community reports more abuse on Twitter since Musk takeover: survey." The Hill, 9 February 2023. thehill. com/policy/technology/3851147-lgbtq-community-reports-more- abuse-on-twitter-since-musk- takeover- survey, retrieved 17 March 2023. Spring, Marianna, op. cit.

4. DEACTIVATED ACCOUNTS

Table 3: October baseline metrics versus deactivated accounts

METRIC	OCTOBER 2022	DEACTIVATED AS OF FEBRUARY 2023	COMPARISON
# of accounts (n)	110,407	7,520	6.93% of total
Median account age	1,531 days	368 days	↓ 1,163 days
Median friends	474	243	↓ 231
Median followers	302	83	219
Median tweets/day, last 200 tweets	22.22	38.8	↑ 16.58
Median tweets/day, lifetime	9.38	9.63	↑ 0.25
Median # of tweets, lifetime	10,724	3,046	↓ 7,678

ELON MUSK'S ACQUISITION of Twitter was characterised by dramatic changes in virtually every aspect of the platform's operations and policies, including draconian cuts to the company's staffing, with a particular focus on the teams responsible for trust and safety, including all forms of content moderation.[20] Despite policy change announcements and staffing reductions, a large number of accounts were suspended during the period of this study, while others voluntarily deleted their accounts and left the platform.

20 Kurt Wagner (with Bloomberg News). "Twitter cuts workers addressing hate speech and trust and safety as Elon Musk's chaotic revamp continues." Fortune, January 7, 2023. fortune.com/2023/01/07/twitter-cuts-workers-hate-speech-trust-safety-elon-musk-revamp, retrieved September 7, 2023; "Days before the midterms, Twitter lays off employees who fight misinformation." NBC News, 4 November 2022. www.nbcnews.com/tech/ tech-news/twitter-disbands-trust-safety-council-rcna61400, retrieved September 7, 2023; Jason Abbruzzese. "Twitter disbands its Trust and Safety Council." NBC News, 12 December 2022. www.nbcnews.com/tech/tech-news/twitter-disbands-trust-safety-council-rcna61400, retrieved 17 March 2023.

As of February 2023, almost 7 per cent of the October dataset (7,520 accounts collected between 10 October and 26 October 2022) had been deactivated, through either suspension or self-deletion. Account status was determined by polling the Twitter user ID number (UID), which was presumed to stay the same for an individual user even when other aspects of the account changed. As noted below, a small number of exceptions to this presumption were discovered when the dataset was interrogated.

In addition to the total deactivations noted above, 15 accounts were labelled 'temporarily unavailable' – a designation of temporary or conditional suspension, usually pending the account holder's decision to remove tweets that violated Twitter's media policy. Another 19 accounts returned unspecified status errors when checked, probably due to glitches in the Twitter API.

I manually examined a random sample of 350 accounts from the deactivated dataset and found that 57 per cent had been clearly labelled as suspended, while 39 per cent were labelled "This account doesn't exist". The latter designation is typically understood to apply to accounts that self-deleted, but it could also indicate an account that was renamed prior to being suspended. The percentage of suspended accounts is probably higher than 57 per cent, perhaps significantly, but is probably not lower.

Glitches in Twitter's infrastructure also affected some of these determinations. Fifteen deactivated accounts, about 4 per cent of the random sample, showed anomalies. Of these:

- Five accounts were still active and consistent with all data collected in October, probably reflecting API errors made during collection, which would put the overall API error rate for the dataset at around 1.4 per cent.

- Five accounts were active with their original handles and displayed identical or similar profile information (such as name, location and bio). However, the accounts had been assigned different UIDs and were labelled as having been 'created' after the October collection period. Previous tweets were apparently deleted. No conclusive explanation for these accounts could be

determined. One possibility is that they self-deleted and then reactivated within 30 days, an option that is available to users who delete their accounts. Another possibility is that the accounts were suspended and later reinstated with a new UID.

- Five accounts showed that the original handle was still active, but they had new UIDs, entirely different profile information and a creation date set after the October collection. These could also have been reactivations after self-deletion, but it is possible that entirely new users were able to register new accounts using the old handles. Previous Twitter policies discouraged or forbade the reuse of handles, but recent news reports suggest that those policies have been under review.[21]

Deactivated accounts generally had less reach than the baseline, with almost half the number of friends and less than a third of the followers. The median age of a deactivated account was 368 days, less than a quarter the age of the baseline.

The deactivated dataset tweeted more frequently than the baseline: 38.80 tweets per day (last-200-tweets method) compared with 22.22. Over 24 per cent of deactivated accounts tweeted more than 200 times per day, compared with 15.4 per cent of the October baseline dataset. Deactivated accounts replied at the same rate as the baseline, but they retweeted 26.85 per cent less, possibly indicating that they were tweeting spam or an automated message.

Median tweets per day, lifetime method, were much closer to the baseline than the last-200 method, suggesting that some deactivated accounts may have sharply increased the frequency of their tweeting shortly before being suspended. Further research might shed light

21 Ryan Mac and Kate Conger. "Twitter Said to Consider Selling User Names to Boost Revenue." New York Times, 11 January 2023. www. nytimes. com/2023/01/11/technology/twitter-user-names-elon-musk.html, retrieved 17 March 2023.

on whether this can be explained by behavioural cues, or by some other cause, such as hackers or spammers taking control of dormant accounts and converting them into bots.

More than 95 per cent of deactivated accounts had a custom profile picture, only slightly less than the baseline. This notably contradicted conventional wisdom which holds that the absence of a custom picture may signal the presence of a bot. Some deactivated accounts had a profile bio that was identical to at least one other account in the network – 117 exact duplicates out of the 5,919 accounts that included at least one character in the bio field. A much larger number had profile information that differed from other accounts by only a handful of characters.

The profiles of deactivated users skewed towards various kinds of sexual content and get-rich-quick schemes. These accounts also showed signs of deceptive or automatic activity; for the most part, it was unclear whether accounts were suspended on the basis of the content described below, or for automated or spammy activity, or whether they self-deactivated.

One cluster of deactivated accounts with profiles in Arabic advertised 'massage services available' in Saudi Arabia, although the accounts did send some tweets in English. At least 63 profiles included the word 'massage' in Arabic, including 55 with identical profile bios and others nearly identical – indicators of bot or spam activity. Another cluster of near-duplicate profiles in Chinese advertised 'gameplay' for men featuring 'models', and yet another cluster included services that 'meet all the needs of men'. Most but not all accounts in the dataset that contained this profile language were suspended.

The clustering of these accounts in the October dataset was an artifact of synchronised tweeting by the accounts during the search process used to create the datasets, rather than necessarily reflecting the overall prevalence of these networks. In other words, accounts in these networks tended to tweet the same content at the same time. By comparison, the January dataset contained only one account with

the Arabic word for 'massage', but it contained multiple accounts promoting sexual content in Chinese, generally using phrases different from those seen in October.

Other deactivated accounts included 'nsfw' (not safe for work) in their profiles, with additional context that suggested sexually explicit content. A medium-sized cluster of deactivated profiles in Chinese contained phrases with clear connections to pornography, including some that might suggest the presence of child sexual abuse material (CSAM). The collection process did not capture media attached to tweets.

After content of a sexual nature, the bulk of the remaining deactivated accounts were implicated in apparent spam or scams. Some accounts with profiles in Chinese advertised the sale of Twitter followers. Other accounts advertised products or content relating to cryptocurrencies (123 accounts), NFTs (155), and gambling.

Many of these accounts showed signs of bot activity. Some of them may have engaged in phishing or spreading malware. Some deactivated accounts claimed that the user was involved in hacking.

Political content was found among the deactivated accounts – in smaller numbers, but far from none. The name 'Trump' was found in 71 profiles, and 'Biden' in 29. In both cases, the accounts represented a mix of pro- and anti- content. Only 10 accounts included 'COVID' in their profiles, while at least 16 mentioned vaccinations, again featuring a mix of pro- and anti- content.

No strong trends were visible with respect to unambiguously extremist content, although a relatively small number of accounts did advocate for or against various extremist ideologies, such as white supremacy.

The 50 most-followed accounts in the deactivated set appeared to include a mixed bag of topics, including politics (in several countries around the world) and cryptocurrencies. Some users had been suspended, while others appeared to have self-deleted. The largest suspended account was attributed to a Nigerian political activist with more than 500,000 followers. The reasons for its

suspension were not immediately clear. The user appeared to have created a new account (which was active when checked in February), as had others among the top 50.

The top 50 skewed noticeably towards the Global South, including accounts claiming locations in Kenya, South Africa, the Philippines, Uganda and Pakistan, among other places.

At least nine deactivated account profiles cited the threat of genocide against residents of the Tigray region of Ethopia.[22] Other accounts tweeting at extremely high speeds (suggestive of bot activity) about the Tigray genocide were found in the main datasets and had not been deactivated. These are discussed more in Section 7.

Finally, at least 87 deactivated profiles made an unambiguous reference to having been suspended at least once previously, while a number of additional accounts included more ambiguous language to that effect.

22 Matti Pohjonen (2022). "An epistemic proxy war? Popular communication, epistemic contestations and violent conflict in Ethiopia." Popular Communication, 20(3), 236–252; Abdi Latif Dahir and Simon Marks. "Slaughter on Eve of Ethiopian Peace Draws Accusations of War Crimes." New York Times, 1 March 2023. www.nytimes.com/2023/03/01/world/africa/ eritrea- ethiopia-tigray-war.html, retrieved 17 March 2023.

5. REINSTATED ACCOUNTS

Table 4: Reinstated accounts by type, as of February 2023

	COVID DISINFORMATION	ALT-RIGHT	QANON
Accounts previously deactivated	9,158	9,967	25,622
Accounts reinstated as of February 2023	3.98% (365 accounts)	6.64% (662 accounts)	2.34% (600 accounts)
Reinstated accounts that resumed tweeting	3.01% (276 accounts)	1.44% (144 accounts)	1.92% (492 accounts)

SOON AFTER TAKING ownership of Twitter, Elon Musk ordered the reinstatement of an unknown number of accounts that had previously been suspended for tweeting extremist content or disinformation. He described this exercise as an 'amnesty'.[23]

Programmer Travis Brown documented the reinstatement of at least 45,000 accounts as of March 2023.[24] Despite the objectively large size of this tally, other data points suggest that the total number of reinstatements may comprise a relatively small percentage of all previously suspended accounts.

To interrogate this question, I examined several collections of deactivated accounts, organised by the topic that probably resulted in their suspension, which I had identified in the course of previous research.

23 "Musk says granting 'amnesty' to suspended Twitter accounts." Associated Press, 25 November 2023. apnews.com/article/elon-musk-technology-donald-trump-business-misinformation-c60bc41229339eaec5008188fa6d057c, retrieved 17 March 2023.

24 Travis Brown. "Elon Musk's suspension reversals." Github github. com/ travisbrown, retrieved 17 March 2023; Casey Newton and Zoë Schiffer. "Why some tech CEOs are rooting for Musk." Platformer, 28 November 2022. www.platformer.news/p/why-some- tech- ceos-are-rooting-for, retrieved 17 March 2023.

For this report, I examined three large samples of deactivated accounts from networks with the following focus:

- A COVID disinformation campaign active in 2020 (n=9,158)
- A network of people who followed accounts self-identified as alt-right (n=9,967)
- A network of people tweeting QAnon hashtags (n=25,622)

All three networks were heavily targeted for suspension under Twitter's previous management, although the lists do include an unknown number of self-deleted accounts. Some accounts may have been suspended for reasons other than the network activity for which I identified them. However, in the case of the QAnon network, I was able to verify in near real time that almost all the accounts had been suspended during an announced crackdown.[25]

In October and early November of 2022 I polled UIDs to confirm that the accounts on each list were still deactivated. When I resampled the lists in February 2023, I found some evidence of account reinstatements. According to posts from reinstated accounts, users had been notified of their changed status via e-mail. Not all reinstated accounts had resumed tweeting.

As of February 2023:

- **COVID-19:** In a network tweeting COVID disinformation, 4 per cent of previously deactivated accounts had been reinstated, and 3 per cent of previously deactivated accounts had resumed tweeting. Forty accounts were labelled "account is temporarily unavailable because it violates the Twitter Media Policy". Typically, this label is applied to accounts that have been temporarily suspended unless and until they remove a specific tweet or tweets.

25 Kate Conger. "Twitter, in Widening Crackdown, Removes Over 70,000 QAnon Accounts." New York Times, 11 January 2021. www.nytimes.com/2021/01/11/technology/twitter-removes-70000-qanon-accounts.html, retrieved 17 March 2023.

- **Alt-right network:** In a network of people who followed alt-right accounts,[26] 6.6 per cent of previously deactivated accounts were reinstated, and 3 per cent had resumed tweeting. Nearly half of the reinstated accounts were labelled as temporarily unavailable.
- **QAnon:** About 4 per cent of previously deactivated accounts were reinstated, and 3 per cent had resumed tweeting. Five of the reinstated accounts were temporarily unavailable. One reinstated account was deactivated again during the collection and analysis period. A small number of these accounts appeared to have tweeted during the period in which they were presumed to be suspended, suggesting either API errors or that they might have been suspended, reinstated and suspended again sometime prior to Musk's 'amnesty' announcement.

Reinstated accounts from the sources above, and a casual inspection of a large sample of Travis Brown's data, appeared to show that reinstated accounts tweeted much less often than the baseline median account.

Important caveats apply to these assessments:

1. None of the reinstated accounts was originally identified specifically on the basis of active tweeting, which was the criterion for the baseline dataset. This may explain some differences in tweeting patterns.

26 J. M. Berger. "The alt-right Twitter census: Defining and describing the audience for alt-right content on Twitter." VOX-Pol Network of Excellence, 2018. www. voxpol. eu/new-research-report-the-alt-right-twitter-census-by-j-m-berger, retrieved 17 March 2023.

2. Some accounts in the deactivated dataset (detailed in Section 4 of this report) appeared to have been recycled with a new creation date and new UID, even though the account profiles gave information similar or identical to that previously collected (contradicting experts' common understanding that a user ID never changes). Account statuses were checked here by UID, a method that may have missed accounts that came back with different UIDs. A spot check of a larger QAnon dataset suggested that this effect was present but minimal, but additional research would be necessary to make a definitive assessment.

3. Most of the reinstatement datasets examined here are too small to inform strong inferences about a typical reinstated account.

In light of these issues, the findings in this section should be treated cautiously.

6. COMPARISONS

HAVING ESTABLISHED METRICS for typical accounts that tweeted in English at a specific moment in time, for comparison purposes in January and February 2023 I collected data on two new datasets, both of which involve content that has reportedly been subject to more permissive conditions since Musk's takeover of Twitter.

- Promotion of COVID-related misinformation[27]
- Use of the 'groomer' slur, which targets LGBTQIA+ people generally, but which has seen increased usage recently as an epithet specifically targeting trans people[28]

The change in COVID policy was formally announced by Twitter,[29] while increased use of the slur was identified by third-party monitoring.[30] In addition to overseeing the reinstatement of many high-profile accounts sharing toxic or problematic content, Musk

27 Ashley Capoot. "Twitter stops policing Covid misinformation under CEO Elon Musk and reportedly restores 62,000 suspended accounts." CNBC, 29 November 2023. www.cnbc.com/2022/11/29/twitter-stops-policing-covid-19-misinformation-under-ceo-elon-musk.html, retrieved 17 March 2023.

28 Kayla Gogarty. "Anti-LGBTQ hate has increased on Twitter since Elon Musk officially acquired the company." Media Matters, 13 December 2023. www. mediamatters.org/twitter/anti-lgbtq-hate-has-increased-twitter-elon-musk-officially-acquired-company, retrieved 17 March 2023; Richard Spangler. "Anti-LGBTQ 'Groomer' Slur on Twitter Surges Under Musk's Ownership, Report Finds." Variety, 13 December 2022. variety.com/2022/digital/news/anti-lgbtq-groomer-slur-twitter-musk-1235459368, retrieved 17 March 2023.

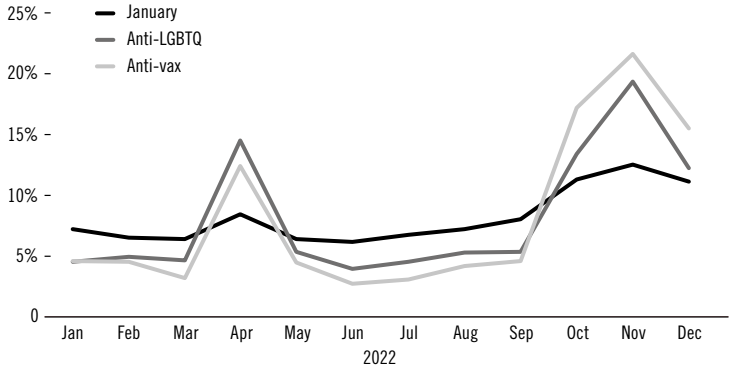
29 "COVID-19 Misinformation." Twitter.com, 23 November 2022. transparency.twitter. com/en/reports/covid19.html#2021-jul-dec, retrieved 17 March 2023.

30 Bond Benton, Yi Luo, Jin-A. Choi, Keith Strudler and Keith Green (2022). "A Preview of Post Amnesty Twitter: Analysis of 'Groomer' on Twitter After the Colorado Springs Shooting." Center for Strategic Communication, School of Communication and Media, Montclair State University. www. montclair.edu/school-of-communication- and-media/wp-content/uploads/sites/20/2022/11/Montclair-State-Study-Post-Amnesty-Twitter-Outlook-and-Spike-in-Grooming-Terms.pdf, retrieved 17 March 2023.

has personally engaged with and amplified both COVID-sceptical content and anti-trans content on Twitter[31] – sometimes even in the same tweet.[32]

Both datasets showed important differences from the baseline, and from each other. Significantly, both datasets showed new account creation taking place at a faster pace than the baseline dataset, with the biggest relative gains corresponding to Musk's acquisition of Twitter. Both datasets also tweeted more often than the baseline.

Figure 3: New account creation, January baseline versus anti-LGBTQIA+ content and anti-vaccination content



31 Mack Lamoureux. "Elon Musk's Twitter Reinstates Anti-Trans Activists on Same Weekend as Club Q Attacked." Vice, 21 November 2022. www.vice.com/en/article/epz8jz/elon-musk-twitter-colorado-shooting-anti-trans-reinstated, retrieved 17 March 2023; Joe Garofoli. "Elon Musk is leaning into transphobia." San Francisco Chronicle, 13 December 2022. www.sfchronicle.com/politics/article/Elon-Musk-is-leaning-into-transphobia-17649577.php, retrieved 17 March 2023.

32 Elon Musk. Tweet, 11 December 2022: "My pronouns are Prosecute/Fauci." twitter. com/elonmusk/status/1601894132573605888?lang=en, retrieved 17 March 2023.

6.1 ACCOUNTS THAT TWEETED ANTI-VACCINATION HASHTAGS

Table 5: Anti-vaccination tweeters compared with the January baseline

METRIC	JANUARY 2023	ANTI-VAX HASHTAG	DIFFERENCE
# of accounts (n)	108,723	10,466	NA
Median age	1,583 days	1027	↓ 556
Median friends	441	563	↑ 563
Median followers	289	331	↑ 42
Median tweets/day, last 200 tweets	18.18	33.17	↑ 14.99
Median tweets/day, lifetime	8.87	12.17	↑ 3.30
Median # of tweets, lifetime	10,404	9,276	↓ 1,128

While its efforts were far from comprehensive, Twitter did take steps to prevent the spread of COVID misinformation prior to the Musk acquisition. As of 23 November 2022, those efforts were formally suspended, with Twitter announcing that it would no longer take enforcement action against vaccine – or other COVID-related misinformation or conspiracy theories.[33]

Anti-vaccination extremists (in the context of COVID-19 and successor viruses) are known to use harassing and threatening language online.[34] Some anti-vax adherents have threatened

33 "COVID-19 Misinformation." Twitter.com, op. cit.

34 Kunihiro Miyazaki, Takayuki Uchiba, Kenji Tanaka and Kazutoshi Sasahara (2022). "Aggressive behaviour of anti-vaxxers and their toxic replies in English and Japanese." Humanities and social sciences communications, 9(1), 1–8.

and/or engaged in violence against health-care workers[35] and supporters of vaccination more generally.[36] Anti-vax beliefs and rhetoric are also increasingly co-incident with other forms of violent extremist belief.[37]

In order to analyse this activity, I collected tweets that used the #diedsuddenly hashtag, as well as the related hashtags #suddenlydied and #vaccinedeaths, which appeared in tweets that also contained #diedsuddenly. The #diedsuddenly hashtag and an online video with the same name are used to promote misinformation that falsely attributes untimely deaths to side-effects of COVID vaccines. A Twitter account promoting the video was created just before Musk took over the platform.[38]

Accounts tweeting the hashtags were identified using Twitonomy's search function in January and February 2023, with full profiles and tweets collected using the Python app between 10 and 11 February 2023.

35 Sharmila Devi. "COVID-19 exacerbates violence against health workers." The Lancet, 396(10252), 658; American Hospital Association. (2022). Fact sheet: Workplace violence and intimidation, and the need for a federal legislative response.

36 For example, Nicole Winfield. "Italy: Police move against violent anti-vaccine activists." Associated Press. 15 November 2021; Ahmed Aboudouh. "'Unparalleled threats': Anti-vaxxer movement threatens a new wave of extremism." The Independent. 19 January 2022. apnews.com/article/coronavirus-pandemic-turin-italy-europe-health-754680bc64a57b30b1701b7ba9faccc3, retrieved 9 September 2023. www.independent.co.uk/news/world/europe/antivaxxer-extremism-violence-plots-b1991710.html, retrieved 9 September 2023.

37 Will Carless. "Far-right extremists look to bring vaccination opponents into their fold." USA Today. 23 January 2022. www.usatoday.com/story/news/nation/2022/01/23/anti-vaccination-ideas-extremists/6599746001, retrieved 9 September 2023.

38 Kaitlyn Tiffany. "Twitter Has No Answers for #DiedSuddenly." The Atlantic, 24 January 2023. www.theatlantic.com/technology/archive/2023/01/died-suddenly-documentary-covid-vaccine-conspiracy-theory/672819, retrieved 17 March 2023.

The most common keywords in the profile bios included truth, freedom, patriot, conservative, Trump, MAGA and Christian. The most common multiword terms included 'truth seeker', 'question everything', 'America first', 'pure blood', 'constitutional conservative', 'American patriot' and 'critical thinker'.

A very small number of users appeared to have shared the anti-vaccination hashtags for the purpose of criticising them.

The anti-vaccination dataset differed from the January baseline in several respects. Accounts that tweeted anti-vaccination hashtags had notably higher median friend and follower counts than the baseline account – 27.66 per cent and 14.53 per cent higher respectively – despite the fact that the median account age was 35.12 per cent lower than accounts at the January baseline.

The lower account age was largely explained by two massive surges in new account creation at a higher pace than at the baseline: a first wave in October 2022, when Musk completed his purchase of Twitter, and a second in November 2022, when Twitter announced that it would stop policing COVID misinformation.

Total lifetime tweets were 10.84 per cent lower than the baseline figure, again reflecting the surge of new accounts, but a typical anti-vaccination account tweeted 33.17 times per day (last-200 method), 82.42 per cent more often than at the January baseline. Of the anti-vaccination dataset, 14.57 per cent tweeted 200 times per day or more (last-200 method), compared with 10.78 per cent of the baseline dataset. However, no anti-vaccination accounts tweeted more than 1,000 times per day (lifetime method).

Anti-vaccination accounts replied to tweets 11.11 per cent less often than the baseline (in their last 200 tweets), but they retweeted 46.39 per cent more. The account that received the most retweets, by a wide margin, was associated with the anti-vaccination video that shared the hashtag's name.

6.2 ACCOUNTS THAT TWEETED AN ANTI-LGBTQIA+ SLUR

Table 6: Accounts that tweeted an anti-LGBTQIA+ slur compared with the January baseline

METRIC	JANUARY 2023	ANTI-LGBTQIA + SLUR	DIFFERENCE
# of accounts (n)	108,723	13,377	NA
Median age	1,583 days	1,724	↑ 141
Median friends	441	568	↑ 127
Median followers	289	281	↓ 8
Median tweets/day, last 200 tweets	18.18	28.29	↑ 10.11
Median tweets/day, lifetime	8.87	11.69	↑ 2.82
Median # of tweets, lifetime	10,404	12,627	↑ 2,223

The misuse of 'groomer' in online discourse – when the word is used, wrongfully, to associate LGBTQIA+ people with paedophilia – has increased in recent years, especially since 2019, and especially in the context of hate speech against trans people.

Use of the term has surged on Twitter in a real-world context of sharply increasing extremist violence targeting transgender and non-binary people.[39] Use of the slur has increased on Twitter during Musk's tenure.[40] Shortly after the period covered in this study, Twitter formally revoked policies banning hateful conduct towards transgender users.[41]

Tweets that included the slur were collected at staggered times between 26 January and 5 February 2023, for a total of 13,377 accounts. The last 200 tweets from each user were collected on 5 February.[42]

The most common keywords in user profiles included proud, fan, conservative, mom, lover, Trump, pronoun, wife, truth, patriot and MAGA. Top multiword terms included 'free speech', 'animal lover', 'America first', 'common sense' and 'Trump supporter'.

The word 'Democrat' appeared in 283 profiles, compared with 188 profiles containing 'Republican' or 'GOP'. 'Trump' appeared in 386 profiles, while 223 profiles contained the text string 'trans',

39 California state government report. "Rates of physical and sexual violence higher against non-binary and trans people." 7 September 2023. www. openaccessgovernment.org/rates-of-physical-and-sexual-violence-higher-against-non-binary-and-trans-people/166035, retrieved 9 September 2023. Hannah Schoenbaum. "Report says at least 32 transgender people were killed in the U.S. in 2022." Associated Press. 16 November 2022. www.pbs.org/newshour/nation/report-says-at-least-32-transgender-people-were-killed-in-the-u-s-in-2022, retrieved 9 September 2023. "Interest over time for 'groomer – United States, 1/25/13–2/25/23'." Google Trends. trends.google. com/trends/explore/TIMESERIES/1677352800?hl=en-US&tz=300&date=2013-01-25+2023-02-25&geo=US&q=groomer&sni=3, retrieved 25 February 2023; Aoife Gallagher and Tim Squirrell, "The 'Groomer' Slur." Institute for Strategic Dialogue. 16 January 2023. www.isdglobal.org/explainers/the- groomer-slur, retrieved 25 February 2023.

40 Kayla Gogarty. op. cit.; Bond Benton, Yi Luo, Jin-A. Choi, Keith Strudler and Keith Green. "A Preview of Post-Amnesty Twitter." op. cit.

41 Angela Yang. "Twitter quietly changes its hateful conduct policy to remove standing protections for its transgender users." NBC News. 18 April 2023. www. nbcnews.com/tech/twitter-changes-hateful-conduct-policy-rcna80338, retrieved 9 September 2023.

42 Short for 'Not Safe For Work', often used to flag graphic content.

with 8 including the hashtag #transrightsarehumanrights, while 255 contained 'pronoun' or 'pronouns'. Most of these terms were used in both positive and negative contexts. Some users tweeted the search term to criticise those who used it to defame trans people, but most of the dataset was anti-trans in orientation. The most influential and most-followed accounts included several well-known anti-trans accounts that drive substantial engagement on Twitter.[43]

Median metrics for this dataset differed considerably from those of the baseline user. Unlike the younger accounts tweeting anti-vaccination content, accounts that tweeted the slur were almost 9 per cent older than the baseline group collected in January 2023, with a median age of 1,724 days compared with 1,583.

Scrutiny of the data suggested that there was no single explanation for the difference from the baseline. Creation date trends in the slur dataset tracked very closely with the baseline dataset, with accounts being added at a slightly faster pace than the baseline across several years, including a spike in April that substantially overperformed. The difference in age between this and the anti-vaccination dataset, whose median was younger than the baseline, is probably attributable to the fact that anti-vaccination content was aggressively policed by Twitter prior to Musk's takeover, as opposed to anti-trans content, which operated in a much more permissive environment.

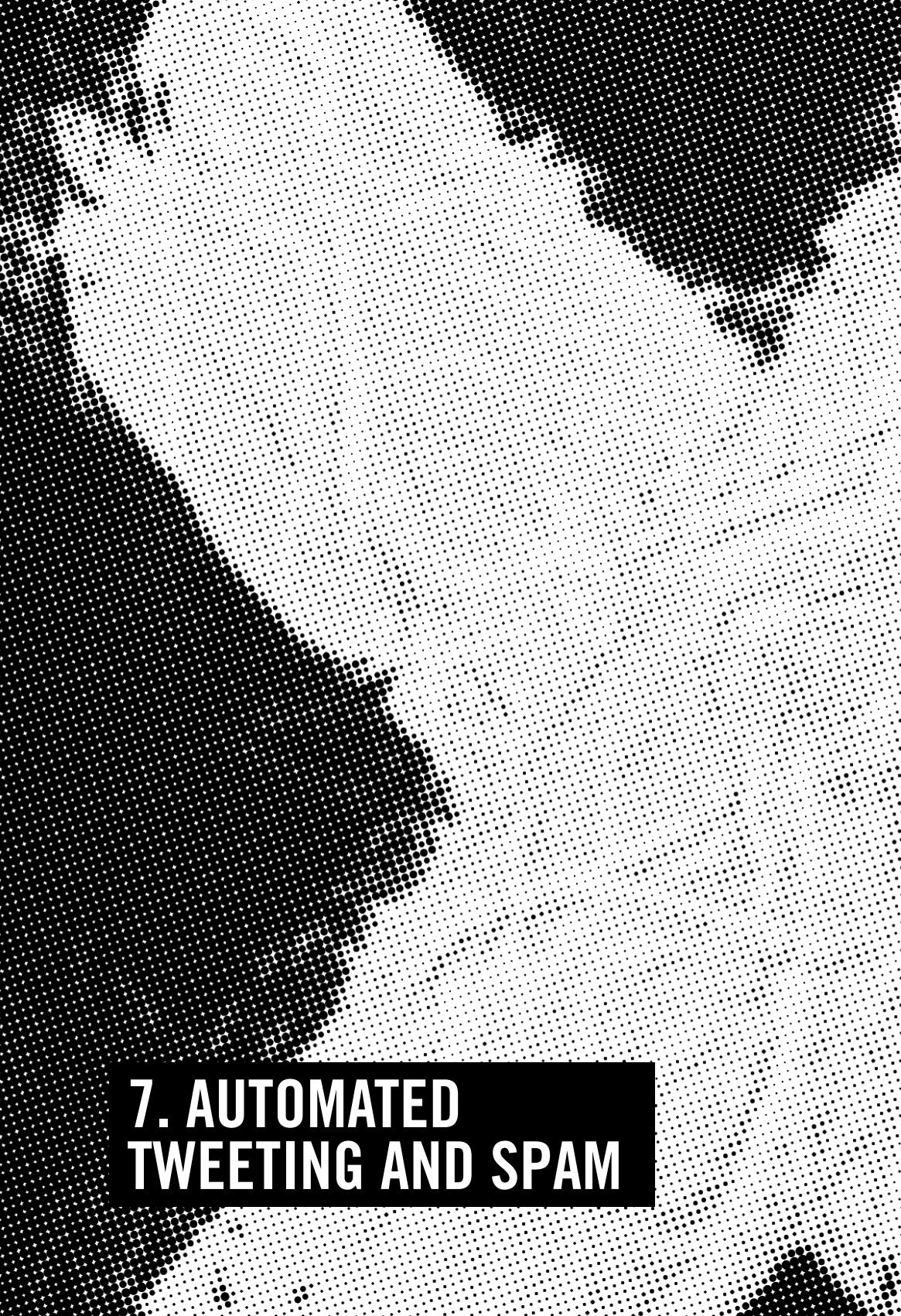
Slur-dataset accounts followed almost 29 per cent more accounts than the baseline but had slightly fewer followers: a median of 281 compared with the baseline's 289.

Number of tweets per day was much higher than the baseline: 28.29 compared with 18.18 (last 200 method). This number was also higher when calculated on a lifetime basis: 11.69 tweets per day versus the baseline's 8.85. Accounts tweeting more than 200 times per day (last-200 method) came in at 13.76 per cent of the dataset, compared

43 Christopher Wiggins. "Elon Musk Joins Anti-LGBTQ+ Libs of TikTok to Mock Fired Employee." Advocate, 16 November 2022. www.advocate.com/news/2022/11/16/elon-musk-joins-anti-lgbtq-libs-tiktok-mock-fired- employee, retrieved 17 March 2023.

with 10.78 per cent of the baseline dataset. No accounts in the slur dataset tweeted more than 1,000 times per day (lifetime method). Total lifetime tweets were about 20 per cent higher than the baseline.

Slur-dataset accounts replied to other accounts 51.85 per cent more often than the baseline dataset, and they retweeted 23.76 per cent more. This made for a marked contrast with the anti-vaccination dataset, whose accounts retweeted more than the baseline but replied less. The most replied-to account targeted by network users was that of the president of the United States. American politicians and public figures from both parties filled the top 50 list, skewing Republican, which suggests that the replies were fuelled by a mix of argumentation, trolling and/or lobbying campaigns.



7. AUTOMATED TWEETING AND SPAM

Table 7: Automated tweeting over time

METRIC	OCTOBER 2022	DEACTIVATED	JANUARY 2023
# of accounts (n)	110,407	7,539	108,723
200+ tweets per day	15.45% (17,057 accounts)	24.07% (1,815 accounts)	10.78% (11,718 accounts)
1000+ tweets per day	0.12% (131 accounts)	0.21% (16 accounts)	0.10% (110 accounts)
Identical profiles	1.57% (1,455 of of 92,757)	1.98% (117 of 5,918)	1.50% (1,353 of 90,023)

IDENTIFYING AUTOMATED ACTIVITY online has traditionally been a tricky business. Not all accounts that use automation are pure 'bots' that tweet without any human intervention, and the term is often used carelessly in the media and in online discourse. Bots come in many varieties, some 'good' and some 'bad', and their characteristics are constantly evolving, as the creators of bad bots seek to evade detection and suspension.

When Elon Musk discussed acquiring Twitter, he made much of the 'problem' of bots, giving various estimates of their prevalence and vowing to deal with the problem. Prior to the takeover, Twitter estimated that 'spam bots' accounted for 5 per cent of its active-user base, a figure Musk's legal team claimed was as high as 33 per cent prior to the acquisition.[44] One third-party estimate split the difference, with an estimate of 9 per cent to 15 per cent.[45]

44 Assaf Dar. "We Checked Elon Musk's Claims About Twitter Bots; Here's What We Found." CPO Magazine, 7 December 2022. www.cpomagazine.com/cyber-security/we-checked-elon-musks-claims-about-twitter-bots-heres-what-we-found, retrieved 17 March 2023.

45 Onur Varol, Emilio Ferrara, Clayton Davis, Filippo Menczer and Alessandro Flammini. (2017, May). "Online human-bot interactions: Detection, estimation, and characterization." In Proceedings of the international AAAI conference on web and social media (Vol. 11, No. 1, pp. 280–289).

Many factors complicate these questions, but we can extract some findings from the data in terms of the prevalence of bots and the impact of Twitter policy on bot operations.

In the December tweet database, each tweet returned by the API included a 'source' field, which included the name of the client purportedly used to send the tweet and a URL purportedly linking to that client's website. The source is intended to show how the tweet was sent, for instance by a mobile phone client ('Twitter for Android') or by clicking a 'share' link in an app ('Instagram').

The source field makes it possible to identify some automated activity definitively, but not to rule out additional activity. The field is not a reliable narrator, as can be seen by numerous mismatches between the client name and URL. Some clients that were not Twitter used Twitter.com in the source field, while some clients claimed to be Twitter apps but linked to non-Twitter domains. Given this, it is reasonable to assume that some non-Twitter bots listed a Twitter URL and the Twitter name.

About 5 per cent of all tweets were clearly attributed to a non-Twitter client, with 2,115 non-Twitter clients identified in the source field of tweets, including 212 clients that contained the text string 'bot' in the URL or description, sometimes as branding rather than strictly accurate description. 'Bot' clients sent 277,830 tweets, or 1.4 per cent of the full dataset. Many other tweets were attributed to well-known automation or social media management services such as IFTTT, SocialFlow and Sprout Social.

In the March resample dataset, the number of unambiguous non-Twitter clients dropped to 1,722. The number of clients with 'bot' in the name dropped to 135, sending a total of 236,150 tweets, or 1.2 per cent of the full dataset. This was the first of many data points to suggest that Twitter had begun cracking down on bots.

In addition to the source field, many outlinks pointed to services known to automate the posting of content such as weather, horoscopes or alerts when a user gains or loses followers. These

so-called 'consumption bots',[46] while sometimes legitimate in and of themselves, are frequently seen in networks that contain less-innocuous automated activity, such as propaganda and disinformation.[47]

The most-linked domain in the dataset was bit.ly, which allows users to track who clicks on their links. All of the bit.ly links in the top 20 outlinks pointed to a bot that tweets personalised horoscopes from the accounts of users who sign up for the service. The dataset also contained source data and outlinks pointing to other apps that automate or schedule tweets, such as Buffer, Sprout Social, Social Flow and IFTTT.com.

Additional clusters of bots could be identified by their very fast pace of tweeting. This signal can be misleading when applied to individual accounts, as some human users do tweet at prodigious rates. Users engaged in online harassment or spreading conspiracy theories, for instance, may organically tweet faster than the baseline, as seen in Section 6. Bots typically tweet faster than the baseline, but not all fast-paced tweeters are bots, and not all bots are fast. Nevertheless, the metric is useful in aggregate and at the absolute highest thresholds.

For detecting unambiguously automated activity, the lifetime method for determining tweets per day was more useful than the last-200 method. For instance, a reporter covering a breaking news story might tweet 200 times in a single day, even if they tweet at a more measured pace under normal circumstances. But it would be extraordinarily unlikely for any given account tweeting more than 1,000 times per day to do so without the assistance of automation.

46 Richard J. Oentaryo, Arinto Murdopo, Philips K. Prasetyo and Ee-Peng Lim, "On profiling bots in social media." Social Informatics: 8th International Conference, SocInfo, November 11–14, 2016. ink. library.smu.edu.sg/cgi/viewcontent.cgi?article=4650&context=sis_research, retrieved 17 March 2023.

47 J. M. Berger. The alt-right Twitter census. op. cit.

In the October dataset, 131 accounts tweeted a thousand times or more per day (lifetime). The most prolific account tweeted mostly in Japanese about an American fried chicken franchise, clocking in at 3,461 tweets per day sent to its 2.5 million followers. One account in the January dataset described itself as a 'test account' and sent seemingly random text strings and links at the staggering pace of almost 17,000 tweets per day – nearly a dozen per minute. Even teams of humans working together around the clock would struggle to match this level of output.

The timeline shows clear reductions in prolific tweeting, consistent with a Twitter crackdown on bots. In October 2022, accounts tweeting more than 200 times per day comprised 15.4 per cent of the dataset, compared with 24.1 per cent of the deactivated dataset and 10.8 per cent of the January 2023 dataset. Accounts tweeting 1,000- plus times per day comprised 0.12 per cent of the dataset in October, 0.21 per cent of deactivated accounts, 0.10 per cent of the January dataset, and 0.09 per cent of the March dataset. A total of 370 accounts in the January dataset had handles that ended with the text string 'bot', down from 437 accounts in the October dataset.

Three of the four most-followed bot accounts in January claimed to enable users to download videos posted to Twitter. The remaining bot tweeted a Taylor Swift lyric every hour. Suspended bots in the deactivated dataset included one that claimed to retweet everything mentioning the hashtag #Web3, one that claimed to retweet anime content, and one that claimed to retweet dog videos.

Taken together, all of these data points suggest that Twitter took action to curtail the use of bots between October and January. Given their typically elevated pace of tweeting, it is possible that the removal of bots fully explains the decline in tweets-per-day between October and January.

Even using the most generous estimates that can be derived from this data, the total number of bots on the platform seems likely to be closer to pre-acquisition Twitter's 5 per cent estimate than to Musk's

33 per cent. However, the tendency for bots to tweet faster than the baseline leaves open a strong possibility that bot-sent tweets as a total of all tweets exceeded 5 per cent, perhaps substantially.

The campaign against bots was still in flux as of this writing, with more dramatic changes expected. In February, Musk announced that Twitter would start charging for access to its API (Application Programming Interface), the access point used by most bots to send tweets. The API pricing move sparked widespread concerns that bad actors (such as spammers or cyber-attackers) would be willing to pay for access to protect their profits, while 'good' bot operators (offering informative or humorous content) might be unwilling or unable to absorb the cost.[48]

Although the pricing scheme had not yet been implemented as of this writing, several prominent consumption bot accounts pre-emptively said that they would cease operation,[49] while others scaled back. Within days of the API announcement, the fried chicken account mentioned above reduced the pace of its tweeting from more than 3,000 to fewer than 5 tweets per day. A different account tweeting in Japanese on behalf of a fast-food franchise abruptly stopped posting thousands of tweets per day, and by February had reduced its pace to 100 or fewer tweets per day.

However, the 'test account' continued tweeting at full speed, as did other clusters suggestive of bot activity. One apparently connected cluster of accounts tweeted thousands of times per day about genocide against residents of Ethiopia's Tigray region, with activity continuing at roughly the same pace after the API pricing announcement. A total of 88 accounts in the January dataset mentioned the Tigray genocide in their profiles, 25 of them tweeting more than 1,000 times per day (lifetime).

48 Ben Cohen. "Elon Musk's New Enemy: An Army of Good Bots." Wall Street Journal, 16 February 2023. www.wsj.com/articles/elon-musk-twitter-bots-api-earthquake-c84d6f11, retrieved 17 March 2023.

49 Tim Fernholz. "Elon Musk killed the good bots." Quartz, 2 February 2023. qz.com/elon-musk-twitter-bots-api-1850065258, retrieved 17 March 2023.

One relatively soft signal for bot activity was the presence of identical or nearly identical profiles in the dataset. Some duplicates were quite minimal and not very diagnostic, for instance a one-character football emoji or an unadorned set of pronouns, but others suggested that multiple accounts had been created for a deceptive purpose.

Of the 90,803 accounts in the January dataset that included at least one character of text in their profile bio, 1.5 per cent (1,353 accounts) had a bio that exactly duplicated at least one other account in the dataset, down slightly from October. By comparison, 1.98 per cent (117 accounts) of the deactivated dataset had duplicate bios, as well as a large number of near duplicates.

It is important to note that the environment may have changed meaningfully since these measurements were taken, and that different methods of bot detection may produce different results. Online observers have documented several instances of low-quality bot activity in April and May, after the period covered in this study.[50] Many of these bots appear to be 'verified' with blue checkmarks.[51]

At least one paper (Hickey, et al.) pre-published in April found little change in bot prevalence under Musk, after assessing bots using a third-party machine learning algorithm. The paper assessed 'hateful' tweets for bot activity, whereas this study looks at a large randomised sample. Additionally, the third-party provider used to assess bot activity worked within API limits that

50 For example, 'Conspirador Norteño' (pseudonym). "The curious case of user one134500 and the most obvious reply spam ever." Conspirador Norteño, 21 May 2023. conspirator0.substack.com/p/the-curious-case-of-user-one134500, retrieved 6 June 2023; 'Conspirador Norteño' (pseudonym). "Inauthentic activity and Twitter Blue verification." Conspirador Norteño, 1 May 2023. conspirator0.substack.com/p/inauthentic-activity-and-twitter, retrieved 6 June 2023.

51 For example, Matthew Gault. "Twitter's T-Shirt Bots Are Undefeated and Verified by Elon Musk." Vice.com, 12 June 2023. www.vice.com/en/article/g5yywm/twitters-t-shirt-bots-are-undefeated-and-verified-by-elon-musk, retrieved 18 June 2023.

restricted the number of accounts that could be assessed.[52] Finally, while the third-party machine-learning algorithm does have some standing in the field, constantly evolving bot behaviours raise some complicated questions about its accuracy at any given time – a problem shared by nearly all Twitter bot detection methodologies, including this author's.[53]

These differences in methodology preclude easy comparison between the two papers. The findings of Hickey et al. may therefore not be inconsistent with those presented in this report. For example, Twitter under Musk's leadership may have targeted mostly low-hanging fruit – bots that unambiguously betray their automated nature through their account names, client names, or volume of tweeting – while neglecting to identify bots that operate in less obvious ways.

52 Daniel Hickey, Matheus Schmitz, Daniel Fessler, Paul E. Smaldino, Goran Muric and Keith Burghardt (2023). "Auditing Elon Musk's Impact on Hate Speech and Bots." arXiv preprint arXiv:2304.04129.

53 Kai-Cheng Yang, Emilio Ferrara and Filippo Menczer (2022). "Botometer 101: Social bot practicum for computational social scientists." Journal of Computational Social Science, 1–18; M. Buğra Torusdağ, Mucahid Kutlu and Ali Aydın Selçuk. (2020, September). "Are we secure from bots? Investigating vulnerabilities of botometer." In 2020 5th (UBMK) (pp. 343–348). IEEE.

8. DISCUSSION

THE STUDY ORIGINALLY set out to dive deeply into questions of what defines a typical Twitter user and how this knowledge can inform our understanding of malicious user activity. Given the rapidly evolving operating environment, however, and multiple concerns about the future viability of Twitter research, I chose to reallocate my time and resources, comparing fewer data points across more datasets.

The final study provides a 30,000-foot view of how Twitter fared during the first few months of Elon Musk's management. Datasets included:

- The October 2022 dataset, collected just before the acquisition.
    - A subset of the October dataset comprising accounts that had been deactivated as of February 2023.
- The January 2023 dataset, which was collected after months of implemented and/or announced changes to Twitter's policy and operating environment.
- Two bad-actor datasets, collected soon after the January one, which included COVID disinformation and anti-LGBTQIA+ extremism.
- Three small datasets of bad-actor accounts reinstated in compliance with Musk-dictated policies.
- A February to March 2023 resampling of the October dataset to examine how surviving users had evolved under the new conditions.

The period of the study saw major changes in site reliability and content moderation policies. But while baseline user metrics changed measurably, and the number of toxic accounts surged, the overall shift seemed less dramatic than narratives around the acquisition might suggest. Most Twitter users kept soldiering on, despite the changes.

Key shifts in the months following the acquisition included the following:

- Between October 2022 and February 2023 the number of accounts followed by the baseline user dropped by 6.96 per cent, and the number following the baseline user dropped by 4.30 per cent. A surge in new accounts between October and December 2022 contributed to the reduction in median friends and followers, as it takes time for new accounts to build their networks.
- Tweets per day (last-200 method) dropped by 18.18 per cent between October and January. Nearly 40 per cent of users in January tweeted less than 10 times per day, compared with 35.6 per cent in October. These declines were driven at least in part by Twitter's enforcement actions, consistent with Musk's stated intention to curtail the activity of bots on the platform. A host of additional data points supported the conclusion that Twitter had engaged in a crackdown on bots, although many bots still remained active as of this writing.
- In the months following Musk's acquisition of the platform on 27 October 2022, Twitter appeared to set an all-time record for new account creation. While new sign-ups began to decline in December, they nevertheless remained at record-high levels. A noteworthy subset of the new accounts identified themselves as conservatives or Trump supporters.
- Almost 7 per cent of the accounts that were active in October had been deactivated by February, with a majority of deactivations resulting from suspensions. Many suspended and deactivated accounts used sex-related terms in their profile bios, while others appeared to reference commercial spam and scams.

Additional findings included:

- Growth in some toxic topic areas appeared to outpace that of the dataset as a whole, including accounts that shared anti-vaccination hashtags and ones that used an anti-LGBTQIA+ slur.

Anti-vaccination accounts were especially responsive to Musk's taking control of Twitter, with sharp spikes on the day after he closed the deal, and on the day when Twitter announced it would no longer enforce its previous policy of banning COVID-related disinformation.

- A review of previously suspended accounts in select networks found that 2 to 6 per cent of previously suspended users had been reinstated. Between 1.5 and 3 per cent of previously suspended accounts resumed tweeting after reinstatement, while the others did not. These results offer a window on reinstatements but may not be generalisable, as the three different networks examined yielded significantly different results, and other potential complications were also observed.

In summary, this report offers a glimpse of Twitter in a time of turmoil. A number of radical changes in policy have been announced, and in many cases implemented, with clearly measurable effects on user activity in most cases.

In these metrics, however, we can see a social network whose attributes include massive complexity and considerable inertia. Short of a technical meltdown, even the most dramatic changes will continue to play out in increments and decrements. Startups steer like cigarette boats; Twitter steers like an ocean liner.

What metrics cannot fully capture is the user experience. What does a 3 per cent drop in followers mean to the typical user? Perhaps not much. What does a 25 per cent drop in retweets mean to someone whose online business depends on reach? Perhaps much more. How does the loss of a third-party Twitter client affect user behaviour? What do users think when their feed contains only tweets from Elon Musk?[54] How does an influx of toxic content affect the health of the ecosystem?

54 Tomás Mier. "Elon's Super Bowl Tweet Flopped, So He Had 80 Engineers Boost His Tweets." Rolling Stone, 14 February 2023. www. rollingstone.com/culture/culture-news/elon-musk-engineers-twitter- engagement-1234680113, retrieved 17 March 2023.

Beneath the relatively stable median scores, changes in tweeting patterns told a story, with an ever-growing proportion of accounts tweeting fewer than 10 times per day alongside a 30.24 per cent drop in accounts tweeting 200 or more times per day. While the reduction in bot activity clearly affected these numbers in a significant way, other factors may also contribute, including site outages and toxicity of dialogue, with the latter broadly signalled by extraordinary growth in the number of users who locked their accounts.

Even an ocean liner must fear icebergs. Despite the slow pace of change in these early months, potentially existential threats loom in the weeks, months, and years to come.

Signs of infrastructure deterioration have been plentiful, in a downward trend that could eventually render the site inoperable.[55] Twitter also has financial liabilities, thanks both to the structure of the acquisition's financing and to erratic business practices under Musk, such as refusing to pay bills submitted by key vendors.[56]

And regulators are watching closely to see whether Twitter's sharply reduced workforce is adequate to meet the company's legal obligations to prevent online harms, including by protecting user privacy and enforcing against copyright violations, extremism, hate speech and child sexual abuse material.[57]

A failure to meet obligations on any of these counts could result in massive fines or even more punitive actions. Numbering among Musk's unpaid vendors is Google Cloud, which provides the infrastructure for what remains of Twitter's Trust and Safety

55 Mike Masnick. "As Twitter Goes Down Yet Again, Report Highlights How Fragile Its Infrastructure Has Become." Techdirt, 2 March 2023. www. techdirt. com/2023/03/02/__trashed, retrieved 18 June 2023.

56 Laura Kolodny. "Elon Musk-led Twitter has been sued by at least six companies for failing to pay bills." CNBC.com, 24 February 2023. www. cnbc.com/2023/02/24/musks-twitter-has-been-sued-by-at-least-six-companies-for-unpaid-bills.html, retrieved 18 June 2023.

57 Michael H. Keller and Kate Conger. "Musk Pledged to Cleanse Twitter of Child Abuse Content. It's Been Rough Going." New York Times, 6 February 2023. www.nytimes.com/2023/02/06/technology/twitter-child- sex-abuse.html, retrieved 19 June 2023.

operation.[58] At the time of this writing, Twitter was apparently trying to bring these functions in-house, even as the days ticked down to the implementation of strict new standards for user safety and privacy under the European Union's Digital Services Act. Regulators had warned repeatedly that Twitter was not ready to meet the 25 August 2023 deadline for full compliance, potentially the biggest foreseeable iceberg on the horizon.[59]

58 David Tuffley. "Twitter is refusing to pay Google for cloud services. Here's why it matters, and what the fallout could be for users." The Conversation, 15 June 2023. theconversation.com/twitter-is-refusing-to-pay-google-for- cloud-services-heres-why-it-matters-and-what-the-fallout-could-be-for-users-207718, retrieved 18 June 2023.

59 Dan Milmo. "'Unprepared' Twitter among tech firms to face tough new EU digital rules." The Guardian, 25 April 2023, www.theguardian.com/technology/2023/apr/25/unprepared-twitter-among-tech-firms-to-face-tough-new-eu-digital-rules, retrieved 18 June 2023.

9. AUTHOR'S REFLECTIONS

IN 2013 I published my first formal study of extremist activity on Twitter, with co-author Bill Strathearn.[60] Ten years later, almost to the day, I completed the data collection for this study, which I expect to be my last.

In the interim I have authored six major studies, three books and countless articles and reports that address how extremists and purveyors of disinformation use Twitter and other social media platforms.[61]

As noted in Section 1 and throughout this study, Twitter has revised its API pricing to eliminate the free API on which I and many other researchers relied. The new pricing is likely to curtail academic and not-for-profit research on the platform dramatically.

Many have noted over the years that Twitter seems to have received more detailed academic and activist scrutiny than other platforms, in no small part because it has been easier to research. As its API was the most open and accessible, commercial products such as NodeXL empowered almost anyone to collect and analyse Twitter data. It can be argued, with some truth, that the price change simply levels the playing field, placing Twitter more on a par with Facebook and other major platforms, where researchers must find creative ways to circumvent API limitations and restrictive terms of service.

Such questions aside, Twitter's API rules made it possible to develop a vast body of open-source research on social networks, thereby enriching our understanding of social media dynamics in general, not just those of Twitter. And while some in the company

60 J.M. Berger and Bill Strathearn "Who matters online: measuring influence, evaluating content and countering violent extremism in online social networks." International Centre for the Study of Radicalisation and Political Violence, 2013. icsr.info/wp-content/uploads/2013/03/ICSR-Report-Who-Matters-Online-Measuring-influence-Evaluating-Content-and-Countering-Violent-Extremism-in-Online-Social-Networks.pdf, retrieved 18 June 2023.

61 icsr.info/wp-content/uploads/2013/03/ICSR-Report-Who-Matters-Online-Measuring-influence-Evaluating-Content-and-Countering-Violent-Extremism-in- Online-Social-Networks.pdf. Op. cit.

might disagree, this independent research was good for Twitter itself, breaking silos and helping to shine a light on problems that might otherwise have been left to fester.

But sunshine is not always an adequate disinfectant. In my first report on Twitter, I examined the followers of some of the most prominent American white supremacists on the platform – a dozen accounts with a grand total of 3,542 followers. The network was working to infiltrate the mainstream of the U.S. Republican Party, but the data showed little evidence that the strategy was working.[62]

Today, the evidence is everywhere. White supremacists and their allied ideologues now have millions of followers on social media. Right-wing extremists have captured a significant, perhaps controlling, interest in the mainstream Republican Party.[63] Twitter's newly permissive policies on moderation have created a much friendlier online environment for extremists.[64] And the account creation data in Section 6 of this report suggests that extremists have seen and answered this invitation in large numbers.

At the same time, Twitter's pro-social functionality has been gutted. Where the platform once played a role supporting pro-democracy activists around the world, it now capitulates to authoritarian regimes abroad, including by facilitating government-mandated crackdowns on political dissent in Turkey and India.[65]

62 Portions of this section are adapted from the author's newsletter. J. M. Berger. "The Last Twitter Show." World Gone Wrong, 11 March 2023. jmberger.substack.com/p/the-last-twitter-show, retrieved 18 June 2023.

63 Jonathan Weisman, and Annie Karni. "G.O.P. Leaders Condemn Lawmakers' Appearance at White Nationalist Conference." New York Times, 28 February 2023. www.nytimes.com/2022/02/28/us/politics/republicans-extremism-marjorie-taylor-greene.html, retrieved 18 June 2023.

64 Charlie Warzel. "Twitter Is a Far-Right Social Network." The Atlantic, 23 May 2023. www.theatlantic.com/technology/archive/2023/05/elon-musk-ron-desantis-2024-twitter/674149, retrieved 18 June 2023.

65 María Antonia Sánchez-Vallejo. "Under Elon Musk, Twitter has approved 83% of censorship requests by authoritarian governments." El País, 24 May 2023. english.elpais.com/international/2023-05-24/under-elon-musk-twitter-has-approved-83-of-censorship-requests-by-authoritarian-governments.html, retrieved 19 June 2023.

In the past, Twitter provided vital resources during the development of news stories and emergencies, such as mass shootings or hurricanes.[66] User verification was a key element supporting this functionality. Instead of validating a user's real identity, Twitter as of this writing was providing a blue checkmark and elevated reach for anyone who paid $8 a month. Even before these changes, users had to exercise judgement to avoid disinformation and misinformation.[67] Now, as blue-check bots and impersonators take advantage of the company's preferential treatment and loosened standards on deceptive or misleading content, pre-existing problems are deepening.[68]

Finally, all of these changes come at a particularly dangerous time, with the proliferation of generative artificial intelligence tools for disinformation[69] and the start of an American presidential campaign that has already availed itself of Twitter's more relaxed standards on COVID-19 misinformation.[70]

66 Alexander Mills, Rui Chen, JinKyu Lee, and H. Raghav Rao. (2009). "Web 2.0 emergency applications: How useful can Twitter be for emergency response?" Journal of Information Privacy and Security, 5(3), 3–26; Mengdie Hu, Shixia Liu, Furu Wei, Yingcai Wu, John Stasko and Kwan-Liu Ma (2012, May). "Breaking news on Twitter." In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 2751–2754).

67 Andrew Weisburd and Clint Watts (2016). How Russia Dominates Your Twitter Feed to Promote Lies (And, Trump, Too). The Daily Beast. www.thedailybeast.com/how- russia-dominates-your-twitter-feed-to-promote-lies-and-trump-too, retrieved 19 June 2023.

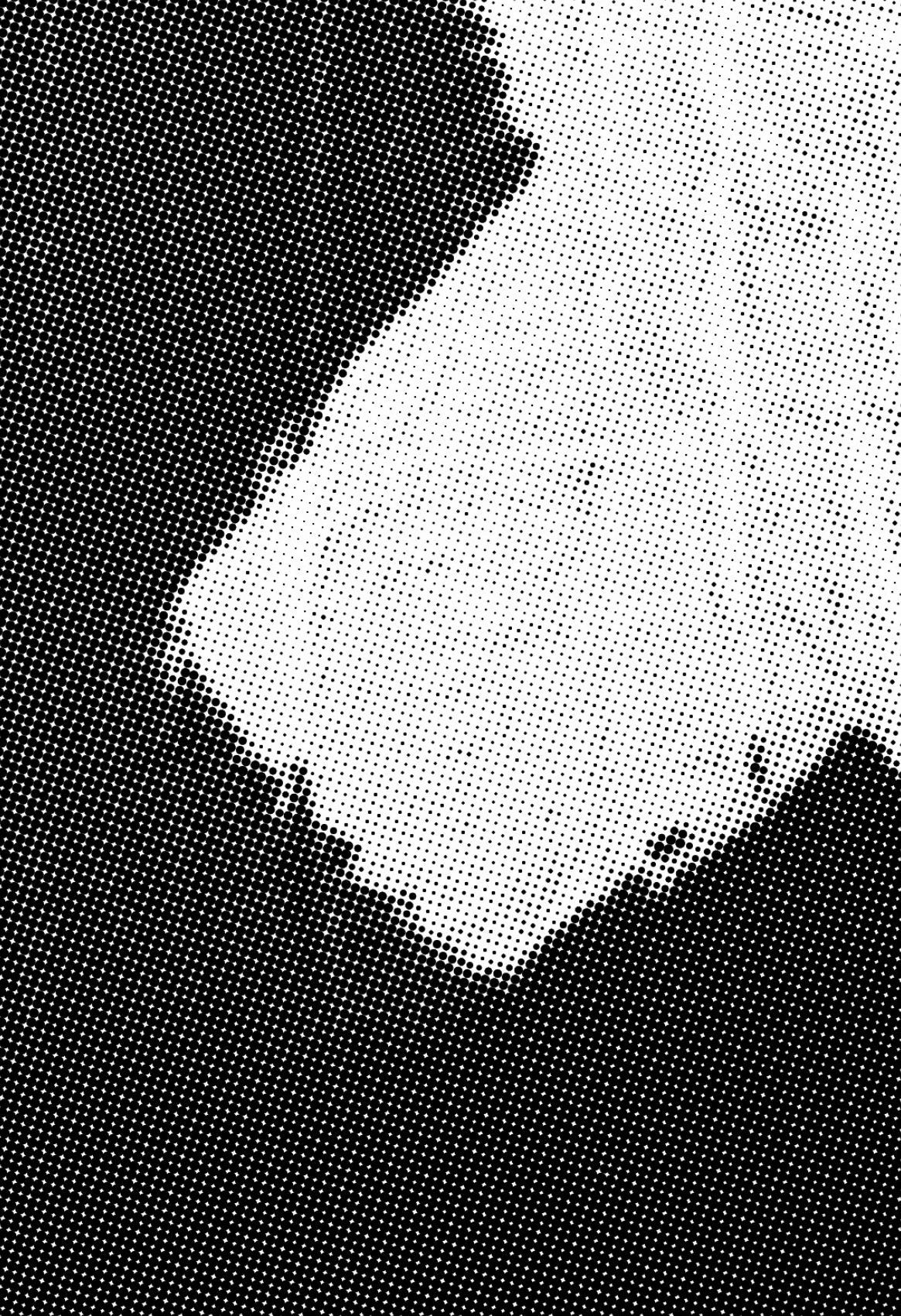
68 Steven Lee Myers, Sheera Frenkel and Tiffany Hsu "Tweets Become Harder to Believe as Labels Change Meaning." New York Times, 26 April 2023. www.nytimes.com/2023/04/26/technology/twitter-verification-problems.html, retrieved 19 June 2023.

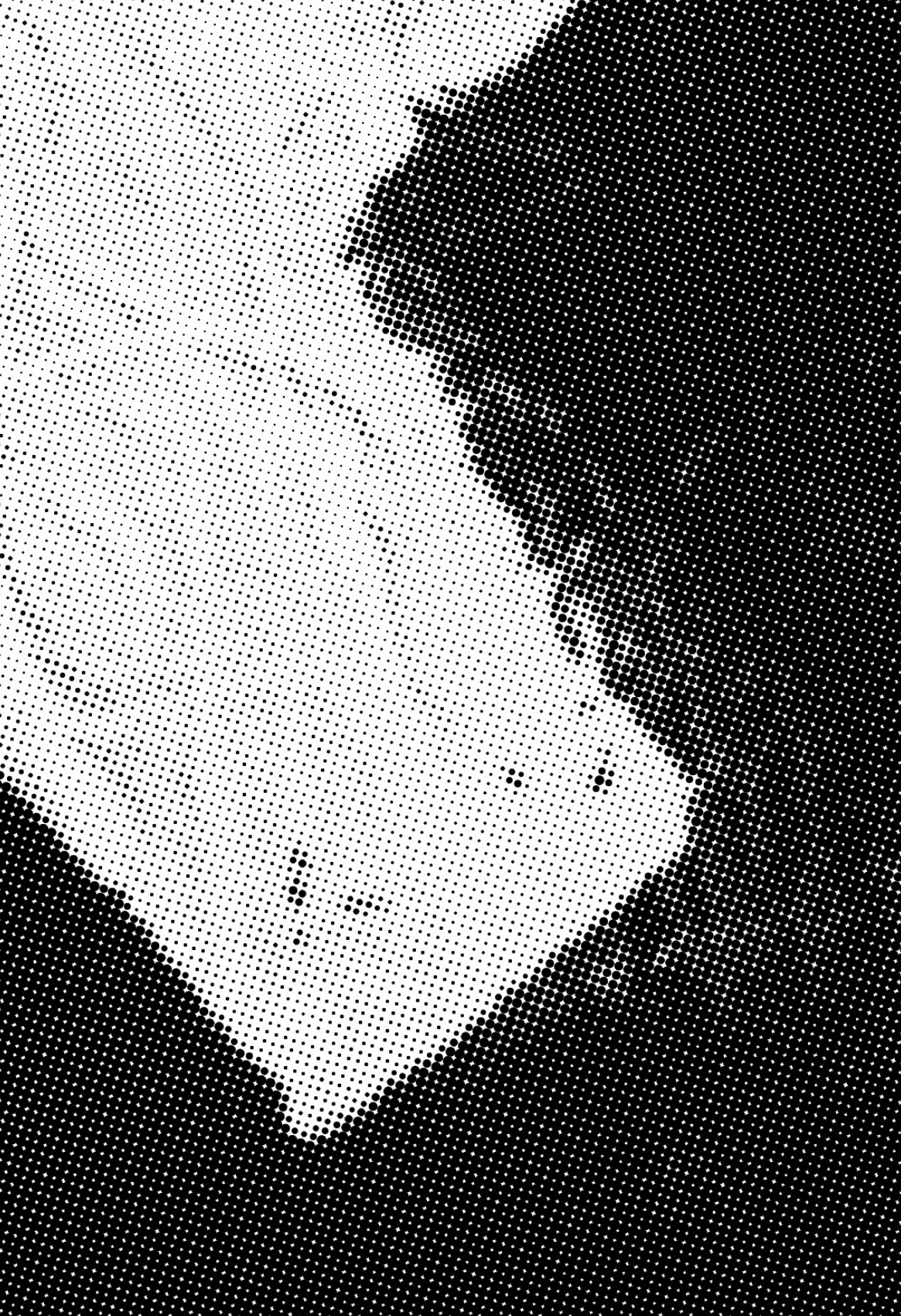
69 Tiffany Hsu and Stuart A. Thompson. "Disinformation Researchers Raise Alarms About A.I. Chatbots." New York Times, 8 February 2023. www. nytimes.com/2023/02/08/technology/ai-chatbots-disinformation.html, retrieved 19 June 2023.

70 Reid J. Epstein, Alyce McFadden and Linda Qiu "Robert Kennedy Jr., With Musk, Pushes Right-Wing Ideas and Misinformation." New York Times, 5 June 2023. www.nytimes.com/2023/06/05/us/politics/robert-f-kennedy-jr-twitter-2024.html, retrieved 19 June 2023.

"The future's uncertain, and the end is always near", as the song goes.[71] It is possible that Twitter will find a way to survive the tempests threatening it, but research on the platform will probably never be the same. This is likely to be the last Twitter Census, at least for this author. Only time will tell whether this study is remembered as the story of a platform beginning to adapt to a period of chaotic change, or an account of its final days.

71 The Doors. "Roadhouse Blues." Morrison Hotel, Elektra Records, 1970, track 1.





APPENDIX: INDEX OF DATASETS

Anti-vaccination dataset: A medium-sized sample of accounts that tweeted hashtags relating to disinformation about COVID vaccines in February 2023. n= 10,466.

Deactivated dataset: A subset of the **October dataset**, consisting of accounts that were deactivated or suspended between October 2022 and February 2023. n=7,539.

December tweets dataset: A database of tweets sent from accounts in the **October dataset**, collected in December 2022. n=19,627,414.

Groomer-slur dataset: A medium-sized sample of accounts that tweeted anti-LGBTQIA+ content in February 2023. n= 13,377.

January dataset: A large random sample of users who tweeted in English in January 2023. n=108,723.

March resample dataset: A dataset resampling the **October dataset** accounts to track user evolution in the February to March 2023 time frame. n=102,422.

October dataset: A large random sample of users who tweeted in English during October 2022. n=110,407.

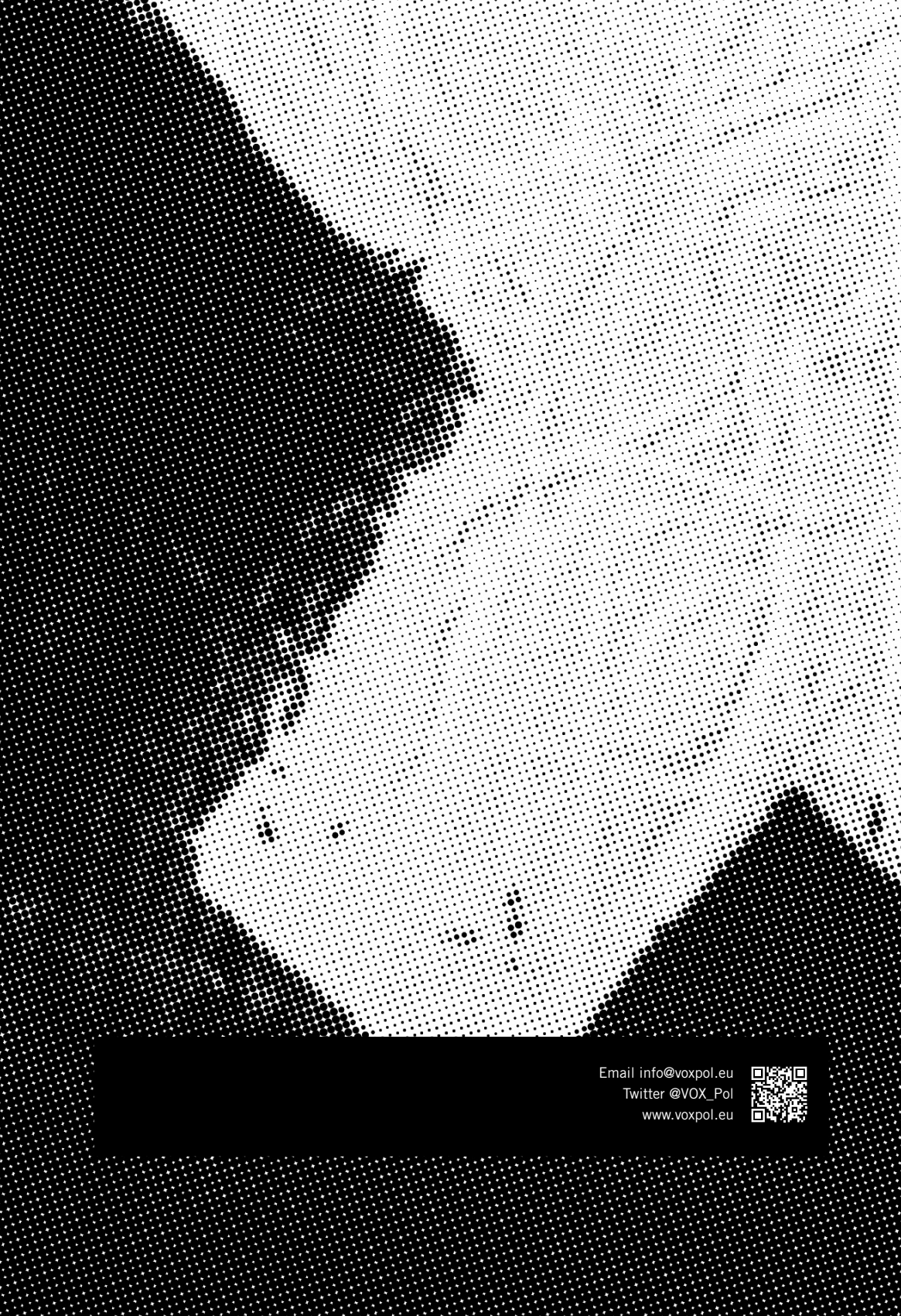
Reinstated datasets: Three smaller datasets containing accounts that had been previously deactivated and then reinstated.

DISCLOSURE

DURING THE PERIOD of this study, Twitter released a number of internal company files to multiple third parties, at the direction of Elon Musk[72] in a possible violation of a federal government order requiring Twitter to take specific steps to protect user privacy.[73] Some of those files claimed to compromise the data privacy of a Twitter account used by the author and others in a 2017 project pertaining to Russian influence networks on Twitter. The data was released by Twitter to writers who used it to impugn the integrity of that project, with Musk's encouragement. The author stands by his work.

72 Will Oremus. "Elon Musk's 'Twitter files' are an exercise in hypocrisy." Washington Post, 16 December 2022. www. washingtonpost.com/technology/2022/12/16/twitter-files- musk-free-speech-hypocrisy/, retrieved 17 March 2023.

73 "FTC probes Twitter data practices after Elon Musk's layoffs." Associated Press, 8 March 2023. www.nbcnews.com/tech/tech-news/ftc-probes-twitter-data-practices-elon-musks-layoffs- rcna74015, retrieved 17 March 2023.



Email info@voxpol.eu
Twitter @VOX_Pol
www.voxpol.eu