global witness

No ifs, many bots?

Partisan bot-like accounts continue to amplify divisive content on X, generating over 4 billion views since the UK general election was called

July 2024

[Picture credit] character style

Contents

Introduction

At the beginning of July, we detected ten bot-like accounts on X spreading political messages around the UK election with an oversized reach. A few weeks later, we've found more – a set of 45 bot-like accounts that amplified partisan content before the election with even greater visibility.

Despite sampling across users who posted hashtags supporting the four largest political parties (by vote share), the majority of bot-like accounts we found expressed support for Reform UK, while just under a third expressed support for Labour.

The accounts' activity is alarming in part for how far it travelled: from May 22 until 9pm on July 4*, the 45 bot-like accounts on X collectively produced around 440,000 posts, garnering over 3 billion impressions.[1] In the 2.5 weeks following the election, their roughly 170,000 posts racked up over 1.3 billion impressions.

What's more, a number of the 45 accounts have continued to amplify divisive political content since the election, including climate denial and 'Great Replacement' conspiracy material. Some have also pivoted to respond specifically to emerging global events, such as anti-migrant protests in Ireland, the assassination attempt against Donald Trump, and Joe Biden stepping down from the US presidential election race: responding with racism, gendered disinformation and conspiracies. In several cases, content shared by some of these accounts was reposted by media personalities with more than 250K followers.

The scale of seemingly inauthentic messaging across a range of sensitive topics suggests this small set of bot-like accounts has outsized influence on user feeds, with concerning ramifications for how users navigate news and events.

Bots that are far from benign

As we've reported in the past, social media accounts that use subterfuge and provocation can pose dangers for public understanding and decision-making – notably around elections. Accounts that mask that they are automated and amplify certain content for political purposes harm democracies by impairing users' access to reliable information and disproportionately promoting misleading material. Their use at scale to manipulate public opinion is a clear and growing threat to democracies worldwide.

Uncovering bot activity is important not just for free and fair elections, but for our ability as societies to tackle critical issues such as climate change. Having previously looked for bots by examining engagement with divisive issues, this investigation instead set out to locate suspected bots that had explicitly promoted different UK political parties. From there, it was possible to get a sample of bot-like activity across the political spectrum, and to examine what the accounts did once the election had finished.

The bot-like accounts our study surfaced were not evenly spread across political parties. Almost two thirds (28) of the 45 accounts expressed support explicitly or shared content expressing support for Reform UK, just under a third (14) for Labour, two for the Liberal Democrats and one for the Conservatives. Nor did their content stick to UK politics – but first it's helpful to know how we found them.

Our method to spot suspected bots

We focused on accounts promoting the four political parties that captured the largest vote share in the UK election: Conservative, Labour, Liberal Democrat, and Reform UK. By searching for a hashtag commonly used to express support for each party ahead of the election (#voteconservative, #votelabour, #votelibdem, #votereformuk), we gathered a sample of posts on X between May 22 and July 4. (We divided the time period in two, taking a sample of up to 5,000 posts from May 22 to June 12 and another from June 13 to July 4.) We chose to focus on X because it's a platform where there are a lot of political conversations.

We removed duplications of accounts when more than one post had been captured in our samples, and searched for the evidence that any of the accounts might be bots.

It’s often not possible to know with certainty that an account is part of a coordinated bot network or troll farm. Our approach focused on surfacing accounts which showed evidence of inauthenticity, automation or disproportionate participation in political conversations online.

We examined indicators that we are calling ‘red flags’, such as accounts that post enormous numbers of posts per day, or that have a handle ending in a long string of numbers indicating the account holder used the default account name suggested by X instead of coming up with their own handle name (see appendix for more red flags).

Having one of these red flags alone does not suggest an account might be a bot, but the appearance of multiple flags, including that the account posts at a higher volume than most humans maintain (more than 60 posts per day), creates reasonable suspicion that the account may have a level of automation. In these cases, we then carried out a manual review of each account and removed any with evidence of their identity, personal life or interactions which were very likely to be authentically human. Even with these checks, however, it is not possible to determine with complete certainty that all of the accounts we have identified are bots.

Assessing political affiliation

We determined the political affiliation of each of the accounts by first tallying the number of posts they posted using each of the four “vote” hashtags. We then reviewed each account for support or critique of any of the parties either explicitly or implicitly in its account bio, cover and profile photos, posts or media content. This helped assess instances in which some accounts used more than one hashtag, or participated positively to one hashtag and negatively to another.

Evidence of network interaction

The bot-like accounts also interacted with each other – through re-posting each other’s content or replying to each other. Often these interactions were between accounts of the same political affiliation. This could suggest (though we cannot prove at this stage) deliberate coordination (which is where bot-like accounts re-amplify each other in order to artificially boost the reach of their content).

Divisive political content posted since the election

In the time since the UK election, a number of the accounts have continued to post divisive political content, including conspiracy theories and climate denial. However some have also pivoted to respond specifically to emerging global events, such as anti-migrant protests in Ireland, the assassination attempt against Donald Trump, and Joe Biden stepping down from the US presidential election race. Conspiracy content shared by the bot-like accounts included references to the World Economic forum controlling politicians and setting up a world government. They also amplified assertions that the Great Replacement theory is ‘proved’ by a UN document from 2000.

14 of the accounts – all of which supported Reform UK – shared the hashtag #ClimateScam in the post-election period. This was often used in conjunction with other conspiracy hashtags such as #geoengineering or #billgates, as well as policy hashtags such as #netzerostupidity.

Climate disinformation shared by these accounts including claims that a 'major new study' had shown that CO2 did not affect global warming, and so the 'globalist parasites' had been proven wrong. The 'major new study' in question was a literature review which claimed that 'the officially presented impact of anthropogenic CO2 increase on Earth's climate is merely a hypothesis rather than a substantiated fact', has not used any new data and has received only one citation.

Other posts shared by this set include claims such as 'humanity is under attack' by globalist weather manipulation, that when Bill Gates met with Keir Starmer they discussed 'the climate scam' and they likely also discussed 'individual carbon trackers' and 'depopulation'.

US politics

This set of accounts has an ongoing interest in US politics, and engages particularly around key flashpoints, described below.

Assassination attempt against Donald Trump

After the assassination attempt on July 13th, there was a major spike in engagement from these bot-like accounts, including driving divisive conspiratorial narratives along political lines.

From July 4th to 13th (inclusive), there were 2,119 mentions of Trump within our dataset. From July 14th to 22nd, there were 10,647 mentions.

For instance, some Reform-supporting accounts shared claims that the attempt was orchestrated by top Democrats such as Biden, or by the CIA in collusion with Barack Obama, Hillary Clinton and Mike Pence. Conversely, some Labour-supporting accounts shared claims that Trump had faked or set up the attempt as a 'show', even though a civilian was killed in the attack.

Joe Biden stepping down and endorsement of Kamala Harris

Kamala Harris has previously been the subject of gendered disinformation campaigns targeting her with racist and sexualised tropes, and with a fake photo purporting to show her with Jeffrey Epstein.

After Biden announced he was stepping down from the US presidential election race, accounts started posting about Harris. This included both support and opposition to Harris in partisan hashtags (such as #harris2024 and #trumpvance2024), and also echoed some of the same tropes that had been previously used against Harris.

These messages included questioning Harris' ethnicity, sexualised comments linking to her past relationship with Willie Brown and derogatory sexualised jokes.

The use of sexualised and racist tropes, and critiquing women’s physicality, are known strategies of gendered disinformation campaigns, which seek to undermine and degrade women on the basis of their identity.[2]

Some of the accounts, however, did also call out disinformation targeting Harris, such as amplifying posts calling out that the alleged photo of Harris with Epstein had been digitally altered.

Anti-migrant narratives

Numerous accounts lit up with posts attacking immigration in the context of an incident in Coolock, Ireland, in which a protest against the use of a factory for accommodation for asylum seekers resulted in fires at the site. Posts promoted the hashtag ‘#IrelandbelongstotheIrish’, encouraged people to join the protests, referenced conspiratorial tropes of ‘population replacement’, or claimed that there was a government plot to import ‘thousands of criminals’ and create chaos to justify greater social repression. Despite the intense focus on a specific local protest in Ireland and the use of nativist hashtags, none of these accounts appeared to be located in Ireland, according to social media monitoring software analysis.

Uncertain provenance

While we have found a larger set of bot-like accounts in this investigation, we have not been able to determine who might be behind them, and there is no evidence that UK political parties are paying for, using or promoting the accounts. The fact that 28 accounts appeared for Reform, and 14 appeared for Labour, suggests potential priorities to elevate right-wing views or a possible motivation to amplify conflicting narratives across the political spectrum. The continued posting of divisive political content after the UK election finished, and the accounts’ movement among topics that are popular on X but are outside of British politics, suggests they could have been created by someone interested in sowing wider discord and stoking tensions in the English-speaking world.

Recommendations

Given the ease with which social media platforms (by their very design) can be exploited to drive divisive and harmful content, responsibility for mitigating risks from their platforms to information integrity and democratic processes must lie with them. In the EU, the Digital Services Act affirms this principle.

Major social media platforms recognise the dangers from harmful bots and have policies that ban them. X’s policies state that users may not “artificially amplify […] information or engage in behavior that manipulates or disrupts people’s experience” and that users that violate this policy may have the visibility of their posts limited and, in severe cases, their accounts suspended.

Unfortunately, these policies are not always adequately enforced. We call on X to investigate whether the list of bot-like accounts that we have provided to them violate their policies and to invest more in protecting our democratic discourse from manipulation.

We wrote to X to give them the opportunity to comment on the investigation’s findings, but they did not respond to them.

Appendix

The investigation used the following list of red flags to assess accounts that appear to be bots. In our methodology, the presence of three or more of these red flags creates reasonable suspicion that the account may have a level of automation.

Common qualities of these kinds of accounts include:

- Deliberate amplification of particular messages or narratives (such as prolific posting and focusing on political messaging)

- Deception or inauthentic activity (which suggests a hidden purpose to the account and is necessary to disguise a bot account)

- Coordination between accounts

We focus in particular on red flags indicating *automation* (bot-like accounts), as this is a strong proxy for coordinated inauthentic activity.

- If accounts are automated at scale, that indicates an intention to influence online conversation

- If accounts are not declared as automated but there is evidence they are automated, that signals a level of deception

Red Flags:

The account is a prolific poster and posts in high volume, likely to be aiming to drive online conversations:

- The account has posted more than 60 times a day on average over the studied period

- And if so, the account has posted more than 144 times a day on average over the studied period

- The account posts in multiple languages (has posted more than once in more than 3 languages)

The account posts a low amount of original or high-quality content and predominantly reposts (reposts are easier to automate, and this may generate some following but is unlikely to lead to mass followings):

- The account reposts other accounts' posts more than 90% of the time

- The account posts original posts (not reposts, replies or quote posts) less than 10% of the time

- The account has posted duplicate content of more than 1 word

The profile was able to be set up quickly in order to respond to an event:

- The account's handle ends in a long string (greater than 5) of apparently random numbers, suggestive of a handle being used that was generated automatically by X

- The account was created in the last 50 days

The profile indicates possible deception:

- The account name and handle do not match in a way that is suspicious

The profile has indications of potential coordination with others:

- An account's posts receive a suspiciously consistent number of likes for its posts
- An account's posts receive a suspiciously consistent number of reposts for its posts

Our manual review of accounts then also took into account:

- The profile picture and cover photo of the profile (has a photo been reused from elsewhere? Are the pictures consistent? Are they easily set up using stock or blank images?)
- The bio of the profile (have they shared details of their identity, or other social media profiles?)
- The posts and replies of the profile (are the replies relevant? Do they refer to personal relationships with other users? Could they plausibly have been generated using a generative AI tool?)
- The media of the profile (have they shared a significant number of personal photos, or photos which confirm their identity? Are there high numbers of repeating or irrelevant images used?)

In addition to vetting the accounts for suspicious signs of automation, we also looked at the extent to which they interacted with each other, which can be a sign of coordinated activity.[3] We found 447 interactions (reposts, replies[4] and quote posts) across the group of accounts from when the election was called on May 22nd until July 22nd, two and a half weeks after the election.

The accounts supporting Reform were the most prolific posters: they drove the most interactions with other accounts, and themselves were interacted with by other accounts the most. Across the set of 45 accounts, interactions often took place within groups of accounts sharing the same political affiliation.

Additional notes:

- We used Information Tracer to help with our analysis.
- We do not have any evidence to suggest that any UK political party is paying for, using or promoting bots as part of their election campaigns.
- Figures are estimated, and may vary where accounts/posts have been deleted.

Endnotes

[1] * We used 9pm on July 4th as a cut-off to show activity before the polls closed (generally at 10pm) and the exit poll figures were announced. Definition of impressions: one impression is measured as the appearance of a post on a user's screen, with the total being the number of times a post has been seen by users.

[2] See: Nina Jankowicz, Jillian Hunchak, Alexandra Pavliuc et al., 'Malign Creativity: How Gender, Sex, and Lies are Weaponized Against Women Online' Wilson Center, January 2021, https://www.wilsoncenter.org/publication/malign-creativity-how-gender-sex-and-lies-are-weaponized-against-women-online#:~:text=Gendered%20abuse%20and%20disinformation%20are,total%20amount%20of%20recorded%20instances; Ellen Judson, 'Gendered disinformation: 6 reasons why liberal democracies need to respond to this threat,' Henrich Boll Stiftung, July 9, 2021 https://eu.boell.org/en/2021/07/09/gendered-disinformation-6-reasons-why-liberal-democracies-need-respond-threat#GenderedDisinfo05:~:text=5.%20Gendered%20disinformation,as%3A%5B51%5D

[3] Timothy Graham, Sam Hames, and Elizabeth Alpert, 'The coordination network toolkit: a framework for detecting and analysing coordinated behaviour on social media,' Journal of Computational Social Science, May 11, 2024, https://link.springer.com/article/10.1007/s42001-024-00260-z

[4] This may include 'inherited' replies - such as when a user in our list reposts or replies to a post that is itself a repost or reply to another account in our list.