Social media bots influence political news coverage

Preliminary Draft: Please do not circulate

Manu Singh[*†]

November 17, 2024

Abstract

I examine how fake social media accounts boost politicians' online popularity and this phenomenon's subsequent spillover on traditional news coverage. Using the 'Botometer' algorithm, I assessed the proportion of bot accounts engaging with tweets from 382 U.S. Congress members on Twitter. A policy change to Twitter's API infrastructure in November 2022 was an exogenous shock to the platform that significantly hampered bot functionality. My first-stage analysis demonstrated that this policy change only affected high-bot-engagement politicians, who saw a substantial decline in followers after November 2022. Placebo comparisons show this decline was not observed in comparable data from Facebook 'likes' or Instagram followers. My second-stage analysis found that after the policy change, high-bot-engagement politicians also saw a decline in coverage in digital news articles and T.V news from December 2022 to February 2024.

[*]I am grateful for comments and feedback from John Marshall, Jacob Shapiro, Andrew Hall, Eunji Kim, Andrew Guess, Donald Green, Gaurav Sood, and Michael Jensen. Participants and discussants at the American Political Science Association 2024, Midwest Political Science Association 2024, Political Economy Seminar at Columbia University, Polmeth 2024, Electoral Integrity Project workshops, Political Communication Pre-conference at University of Pennsylvania, ACM Conference on Economics and Computation at Yale University, and Graduate Student Seminar at Columbia University have provided invaluable feedback. I also acknowledge the support from the Experimental Laboratory for Social Sciences at Columbia University for funding some of this research.

[†]Columbia University

INTRODUCTION

"Twitter is not on the masthead of The New York Times. But Twitter has become its ultimate editor."
—Bari Weiss, Resignation Letter to the New York Times

The volume of online discussion serves as a key indicator of public interest and potential positive opinion, with research showing that high levels of online mentions can predict real-world outcomes. Studies have linked blog mentions to book sales success (Gruhl et al., 2005), and social media mentions to election results in multiple countries (Brito and Adeodato, 2022). Williams and Gulati (2008) and a few others find that 'likes' on Facebook pages and mentions on Twitter predicted political outcomes (Iacus et al., 2012, Franch, 2013, Hong and Nadler, 2012). Moreover, misinformation that receives more social validation—such as 'likes' on social media—is more likely to be believed (Butler, 2024). Research by Smith and Gustafson (2017) indicates that other forms of digital trace data such as Wikipedia page views can supplement traditional polling data. Wikipedia pageviews can predict election results up to 28 weeks before Election Day and significantly improve early-stage forecasting when traditional poll-based predictions are weakest. It is therefore unsurprising retrospectively that Tumasjan et al. (2010) find that not only is Twitter a key platform for political deliberation, but also the number of posts referencing a candidate or party in the pre-election days is correlated with the eventual election outcome.

Next, information-seeking is a crucial aspect of political engagement, with a significant portion of Americans relying on digital platforms for campaign information, especially in the day leading up to an election. Pew research reports that approximately 36 percent of Americans get their campaign information from the Internet, 86 percent rely on digital devices for news, and approximately 59 percent of users on Twitter use it to get news from this platform regularly. Twitter's popularity, accessibility, and reliability as a first-hand source of information about a political candidate make it a useful information-seeking tool. Twitter engagement metrics (follows, likes, reshares, etc.) and posts are valuable indicators of public interest. This concept by itself is not unique and extends to other forms of information-seeking behaviors, including Google search trends and Wikipedia page views. Gómez-Martínez et al. (2022) uses Wikipedia pageviews as an indicator for financial market trends and finds that the volume of queries about an entity is correlated with the evolution of the NASDAQ index. de Silva and Compton (2014) and Mestyán et al. (2013) use Wikipedia views and edits to predict box office success. Multiple researchers also report significant predictive power in Wikipedia trends for forecasting elections and political outcomes (Smith and Gustafson, 2017, Salem and Stephany, 2023, Ciocirdel and Varga, 2016, Yasseri and Bright, 2016).

However, my analysis in Appendix A suggests that Twitter is a trendsetter where political trends first appear. Interest in a candidate, as reflected by their Twitter following, predicts their Wikipedia page views but not vice versa. Research by others such as Giummolè et al. (2013) similarly finds that about 72% of topics appear first as trends on Twitter before Google queries. I wish to emphasize the significant role of social media in shaping public opinion about political candidates and how social media cues and engagement metrics can serve a symbolic purpose (McGregor, 2020). Journalists look to social media to understand public interest; these numbers are used to supplement polls frequently in election forecasting. Political experts actively utilize social media as a tool to gauge real-time public opinion and reactions to events during the campaign period.

However, this media landscape is complicated by the presence of automated accounts (bots) that shape information diffusion and manipulate online political discourse. Ratkiewicz et al. (2011) examined discussions during midterm elections on Twitter and found that social media bots were aggressively expressing support for specific candidates while denigrating others. Similarly, the study by Rossi et al. (2020) and Keller and Klinger (2019) found that political figures in Germany and Finland employed social media bots to amplify follower counts to create an image of popularity and influence. Cook et al. (2014) found bots were used in Australian elections for information dissemination campaigns and inundating newsfeeds to create a 'multiplier' effect for the politician. Bessi and Ferrara (2016) have documented active bot interference in Twitter-based conversations during the 2016 and 2020 U.S. elections.

Moreover, as per Ferrara (2020), bots intensify echo chambers; Users, either because of their preference or algorithmic sorting, engage in selective consumption of content aligned with their preexisting political beliefs and thus are more polarized over time. Abokhodair et al. (2015) dive deeper into the characteristics and behaviors of botnets and find that bots in non-English languages survive on the platform longer and are not necessarily trying to mimic human behavior. These bots were primarily engaged in flooding Twitter feeds with news about the Syrian war. Albadi et al. (2019) find that approximately 11% of hate speech on the Arabic language Twitter could be attributed to bots, with bots participating most actively in highly controversial topics about Israel, Palestine, and Yemen. A famous example of Twitter manipulation in the Gulf by state actors was seen in the aftermath of the murder of the journalist Jamal Khashoggi. A quarter million tweets all just repeating "we all have trust in Mohammed Bin Salman" flooded Twitter to drown out any human conversation[1] (Bell and Coleman, 2018, Leber and Abrahams, 2019, Ignatius, 2018).

Bot interference has been significant in other major political events, with studies showing that bots generated about 20% of tweets in the 2016 U.S. presidential election (Kollanyi et al., 2016) and 33% in the Brexit referendum (Howard and Kollanyi, 2016). Mitt Romney was suspected of having bot followers during his 2012 presidential campaign (Coldewey, 2012), when his follower counts jumped up dramatically. In 2017, the Federal Communications Commission wanted to gather public opinion on repealing net neutrality. Twenty-two million comments were received, about half of which were estimated to be algorithmically generated. Immediately following Russia's official declaration of war on Ukraine on 24 February 2022, the two nations have become embroiled in parallel cyber warfare (Bergengruen, 2023, Bajarin). The use of fake comments,

[1] *"Unfollow enemies of the nation"* was also (re)tweeted over 100,000 times in reference to Al Jazeera's coverage of this incident; *"We have to stand by our leader"* was (re)tweeted 60,000 times all in a span of about 24 hours around 14 October 2018

inflated followers, and engagement metrics, sometimes made using stolen identities or deepfakes, illustrate how online conversation can be manipulated. However, despite bot interference being a well-documented undesirable phenomenon, it has been virtually impossible to establish the causal impact or quantify the influence of such inauthentic activity on real-world outcomes. My research shows that traditional media, such as broadcast television news or online news articles that are typically assumed to be immune to bot-tactics, have hidden vulnerabilities.

Next, social media metrics such as follower counts and engagement rates are valuable currency in politics, directly influencing perceptions of trustworthiness and impact (Castillo et al., 2011). Politicians recognize that higher metrics increase their offline reach and influence (Keller and Kleinen-von Königslöw, 2018, Popa et al., 2020). Moreover, there is a dynamic relationship between politicians and social media conversations. Politicians not only contribute to these conversations but also react to them, often leading to offline consequences (Barberá et al., 2019).

Most importantly, inflated social media numbers could serve as much more than just communication and aid in fundraising and mobilization of motivated individuals (Jungherr, 2016, Parmelee, 2014) by projecting strength, which has the potential to become self-perpetuating. The multi-faceted utility of strong social media metrics illustrates the need for politicians to maintain a positive and robust online presence.

This body of research sets the stage for my paper by framing my key assumption that online popularity and engagement are indirect but a robust measure and predictors of real-life popularity and public opinion. Various entities, including media outlets and political analysts, take cues from social media engagement metrics and draw inferences about public opinion without recognizing that these numbers are highly manipulable through the use of bot accounts.

My paper examines the impact of social media bots on political news coverage and their spillover effects on traditional media. I use the 'Botometer' algorithm to assess bot activity on the Twitter accounts of U.S. Congress members. A key finding is that a change in Twitter's API infrastructure in November 2022 significantly impacted bot functionality, particularly affecting politicians with high bot engagement. This event provided a natural experiment to study the effects of bot activity.

I demonstrate that politicians who relied heavily on bots to inflate their online popularity experienced a notable decline in coverage across both online news articles and television news following the API changes. This suggests artificially boosted social media metrics can influence traditional media coverage decisions. The research highlights how bot activity can create a facade of popularity that extends beyond social media platforms, potentially impacting public perception and media attention.

I also discuss the ethical implications of bot usage in political campaigns and the complexities in regulating such activities. My research challenges some previously held assumptions that only extreme right-wing politicians will likely benefit from bot engagement and the importance of such tactics in close races. While bots may not always spread misinformation directly, they can

Figure 1: Example bot account and some red flags to watch out for

This account has been tweeting almost 500 times per day, every day, for the last nine years. The profile picture or location is not real. Almost every post is a retweet without any original tweets. This account has also posted about 50,000 images and GIFs.

manipulate perceived popularity, which raises questions about the integrity of online political discourse and social media's outsized influence on traditional media coverage.

Theory

What is a bot?

The word 'bot' is an umbrella term encapsulating many kinds of automated online software or scripts. Socially oriented versions of such bots can be programmed to mimic real people on platforms such as Facebook and Twitter (Nimmo, 2018). These bots can engage in political debates and influence public opinion without ethical constraints, potentially spread false narratives, generate misleading data, and impersonate real people to gain credibility. Collectively, the impact is much deeper than just a viral tweet but an ecosystem where genuine human voices are marginalized.

Bots are excellent online manipulators for a couple of reasons. First, the informational nudge by bots is virtually impossible to fact-check as one *like* is one *like*; there is no way for a casual reader to ever know if the person who 'liked' the post is human. This is especially true for popular posts with thousands of such likes.

Next, bots are ideal for *astroturfing*, which is the fake grassroots support of an idea (Ratkiewicz et al., 2011). This centrally coordinated campaign uses participants posing as independent citizens to influence political behavior and electoral outcomes. By manipulating likes, retweets, and other engagement metrics, bots create an illusion of genuine support for political figures or topics (Metaxas and Mustafaraj, 2012, Ratkiewicz et al., 2011). Additionally, bots could be used to spam, drown out dissenting voices, and divert attention from particular conversations.

How bots manipulate platforms

Prior research suggests that when people see information on social media from accounts they agree with, they tend to believe it. Moreover, people re-share information with like-minded friends, and when many people share the same information, the group believes it to be true. However, if social media creates groups of ideologically aligned users, how can messages spread to groups that do not share connections?

This is where social media algorithms play a pivotal role. Social media platforms like Twitter and Facebook utilize algorithms to scrutinize words, phrases, and hashtags, generating a list of topics arranged by popularity. This 'trend list' offers a swift method to examine the most discussed subjects at any given moment. According to Asur et al. (2011) these trending topics can briefly capture widespread attention and influence agenda-setting. Existing online networks supplemented by bot accounts form an ideal foundation for external actors to create a trend and rapidly disseminate a message more efficiently and cost-effectively than any other medium[2]. The algorithm then prioritizes trending or popular topics to be shown to multiple diverse groups. It does not matter who an individual follows; all users see trending or viral tweets as decided by the platform. Moreover, popularity appears to play a bigger role, and mainstream content appears more frequently in users' feeds, leading to a *rich-get-richer* pattern (Napoli and Caplan, 2016, Dujeancourt and Garz, 2023).

Finally, a facade of grassroots popularity can become a self-fulfilling prophecy. The posts can enter a self-reinforcing loop by being on a trending list or the front page; this helps the post get additional views. The algorithm selectively presents this content to new viewers who are likely to find it agreeable. This leads to more public support and is followed by real reposts and shares. All of this contributes to an escalation in a political candidate's perceived online popularity and influence over voters (Persily and Tucker, 2020, Tucker et al., 2018).

[2]When aiming to influence a population through social media trends, the most effective method involves constructing a network of bot accounts designed to post tweets at various intervals, react to specific keywords, or retweet content as a central account instructs. The operational structure of bot networks typically involves a small core of human-operated accounts with large followers who could be dedicated cyber operatives or ideologically aligned enthusiasts. Beneath this core operates a bot network, with bots connecting to each other and the core accounts and repeatedly boosting the core's messages to increase popularity.

Why Traditional media platforms exhibit unforeseen vulnerability to social media manipulation

Social media narratives boosted by bots can gradually shape opinions as propaganda spreads across networks. However, their influence will be limited unless these stories come from a source that an individual finds trustworthy. For example, an individual may see a particular politician trending often and believe they are popular but still may not buy into their messaging or vote for them. This is why it is essential to link the theory of bots to modern journalism practices to understand how bot-boosted popularity can have quantifiable effects even on traditional media.

In a landscape where news can be hyper-personalized and new stories are emerging globally, journalists are also adapting their approaches to compete with alternative and new news sources. As social media platforms such as Twitter become news and information outlets, journalists incorporate it into their strategies to disseminate stories and gather information promptly. An Indiana University School of Journalism study reveals that one of the primary uses of social media by journalists is to stay updated on breaking news. Moreover, mainstream journalists consider tweets a valid source, particularly without more reliable or verified sources. However, the downside is that if a group manipulates a trending topic on Twitter and it catches the attention of mainstream media journalists, they might inadvertently lend credibility to false narratives. Similarly, a politician using bots to appear more popular to voters will also appear popular to a journalist if they are seen on the trending list often and can thus translate online popularity to offline popularity.

Popular topics on social media are demonstrated to influence traditional media agendas by shaping the narrative and news coverage in the traditional media (Wihbey, 2019). Research by Weaver et al. (2019) finds that 59.8% of journalists surveyed in 2018 used social media to find ideas for stories, 56.2% used it to find additional information, and 54.1% used it for additional sources. Similar surveys by the Pew Research Center in 2022 suggest the trend is inching upward. Among journalists who use social media for their jobs, nearly 87% say it has a very or somewhat positive impact on their ability to promote stories, about 80% say it helps them to connect with their audience, and 80% find sources for stories. Most importantly, approximately three-quarters of journalists who use social media find it helpful in identifying stories that should be covered (Gottfried and Liedke, 2022).

Although journalists see the benefits of social media in their line of work, a significant concern is that this new information environment builds a workplace where accuracy is regularly sacrificed for speed as smaller teams are pushed to develop fast news cycles and push stories on multiple platforms. Moreover, user engagement has become a top priority on news channels, which significantly influences editorial decisions and topic coverage[3]. News organizations increasingly rely on metrics such as likes, shares, and comments to gauge audience interest and tailor their content accordingly (Myllylahti, 2020, Lischka and Garz, 2023, Dodds et al., 2023). High-engagement topics often receive more prominence regardless of their traditional news value. While this approach can

[3]Bari Weiss's resignation letter referenced at the beginning of the paper illustrates the frustration of one journalist in how news is prioritized and presented.

make content more appealing to audiences, it also raises concerns about the potential sacrifice of journalistic integrity and the risk of echo chambers forming as news outlets cater to what generates the most engagement rather than what might be most important or informative.

DATA SOURCES

Bot percentage on politicians' pages

In this paper, I use Observatory on Social Media's Botometer tool designed by researchers at Indiana University to assess the likelihood of an account being a bot. The Botometer algorithm uses supervised machine-learning to analyze network characteristics, temporal activity patterns, user metadata, linguistic cues, and sentiment analysis to generate a probability score of whether an account has more human or bot-like characteristics[4]

To create politician-level measurements of bot activity, I scraped all of the tweets in the two weeks before and two weeks after the November 8, 2022, election cycle for all 535 Unites States Congress members[5]. Then, for every tweet, I scraped the Twitter users who had interacted with the tweet by either liking or retweeting the original tweet by the politician. A random sample of this subset was tested using the Botometer API.

There are no rules on what an appropriate threshold is to classify bots. Researchers have employed various thresholds in their studies. For instance, some have used 2.5 (Vosoughi et al., 2018, Shao et al., 2018), while others have opted for 3.5 (Grinberg et al., 2019) or 4.0 (Broniatowski et al., 2018). Yang et al. (2022) suggest conducting analyses with multiple threshold values to assess the robustness of findings. It's worth noting that Botometer versions have evolved, with both detection algorithms and bots becoming increasingly sophisticated. In this analysis, a cutoff of 4.0 was chosen, but the appendix presents all major results for cutoffs ranging from 2.5 to 4.5. As expected, the results are most pronounced when using the highest cutoff, which has the smallest false positive rate. The paper presents the more moderate results using the 4.0 cutoff, and readers can consider the results for higher cutoffs as the upper bound of this analysis. This approach provides a balanced view of the findings while acknowledging the potential range of outcomes based on different threshold selections.

The average rate of bot activity on politicians' Twitter profiles is about 11% (with a median of 10%). This means that about 10% of users who have *interacted* with the politician are bots[6]. I have

[4]The popular version of the Botometer algorithm called *Botometer Lite* is capable of analyzing a very large number of accounts quickly with a shallow examination. For this paper, I have instead chosen to examine fewer (random sample per politician) accounts using the full Botometer analysis in favor of accuracy over speed

[5]The list was taken from the U.S House of Representative Press Gallery and UC San Diego'sTwitter Accounts: Senators websites

[6]These numbers are slightly lower than what other researchers have found because my metric is constructed with a slightly different methodology. Most prior research usually tests accounts that have been collected for a particular topic or a hashtag; in my paper, however, I systematically collect every account that likes or retweets a politician within a specified timeframe and test a random sample of these accounts. I also thoroughly examine a smaller set of accounts using the original Botometer algorithm instead of Botometer Lite

reliable estimates of about 382 politicians. This number is smaller than the number of House and Senate congress members because some have inactive accounts or minimal online activity.

Additionally, I have dropped accounts for whom I could not check over 50 unique accounts that had retweeted the politician. However, this should not be a source of bias because such politicians are not the target of my theoretical predictions. I aim to compare politicians active on social media with and without bots. Nevertheless, all results presented in this paper are robust to including politicians with minimal Twitter presence as 'low-bot users'. Lastly, I acknowledge legitimate concerns about the account-level accuracy of the Botometer API in identifying bot accounts, but my approach mitigates these limitations. By analyzing hundreds of accounts that retweet each politician, I can discern a meaningful signal from noise on a per-politician basis. An aggregate politician-level analysis helps me separate true bot activity from potential noise or false positives that might occur at the individual account levels. My first-stage results also lend credibility to my identification of high and low-bot-engagement politicians.

Table 1: Summary statistics for politicians

	N	Mean	SD	Median	Min	Max
Followers Count	382	313105	1013291	52549	1141	12503380
Tweet Count	382	8275	7069	6742	303	67677
Bot Count	382	637	1982	41	0	15959
Checked Account	382	6204	18389	400	51	139086
Bot Percentage	382	11%	7%	10%	0%	59%

Note: Summary statistics for 382 congress members and senators. Politicians with fewer than 1000 followers during data collection or where fewer than 50 separate retweeting accounts were examined have been excluded to reduce measurement errors. These accounts fall in the category of negligible Twitter users. Results are robust to including these politicians in the low bot usage category.

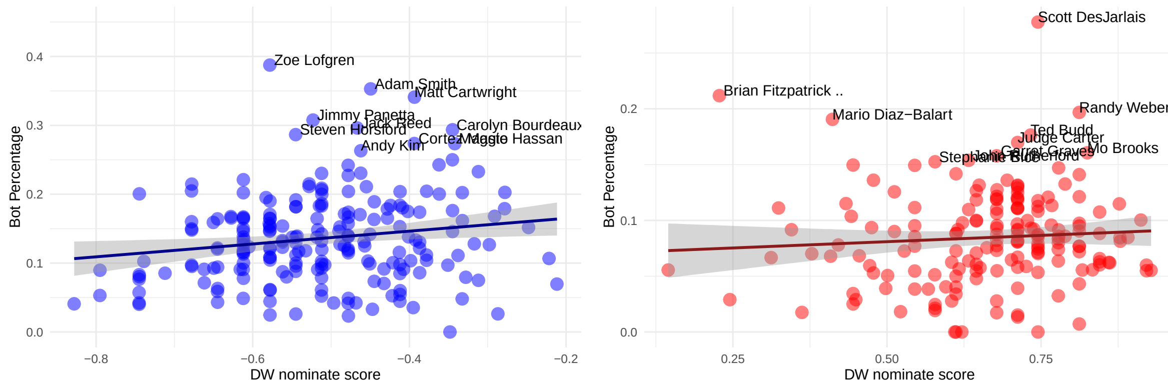
Description of politicians with high bot engagement

Ideology

Politicians in precarious positions, those with low public recognition, or those holding extreme views may be strongly motivated to employ bots for manipulating social media popularity to gain electoral advantages. A previous study by (Silva and Proksch, 2021) found significantly higher bot activity among far-right politicians, particularly those with strong anti-EU rhetoric. Other research indicated that bots generated approximately 20% of tweets with pro-Trump content during the 2016 U.S. presidential election, (Kollanyi et al., 2016), and 33% of tweets in the Brexit referendum (Howard and Kollanyi, 2016).

However, my analysis of local politics challenges previous assumptions about bot usage in political social media strategies. Republican representatives show lower bot engagement, with no correlation to ideological extremity. Surprisingly, Democrats have higher bot engagement, particularly among moderates rather than the far-left. There is a weak but significant relationship between the DW-Nominate score and bot percentage amount democrats.[7]

Figure 2: Relationship between ideology and bot engagement rates



Note: It is important to remember here that there is no evidence that the politicians or affiliates themselves are buying bots. Proving bot origin and purchaser is virtually impossible. My claim is that around the mid-term elections of Nov 2022 the above politicians had bot activity on their twitter feeds.

Close and competitive races

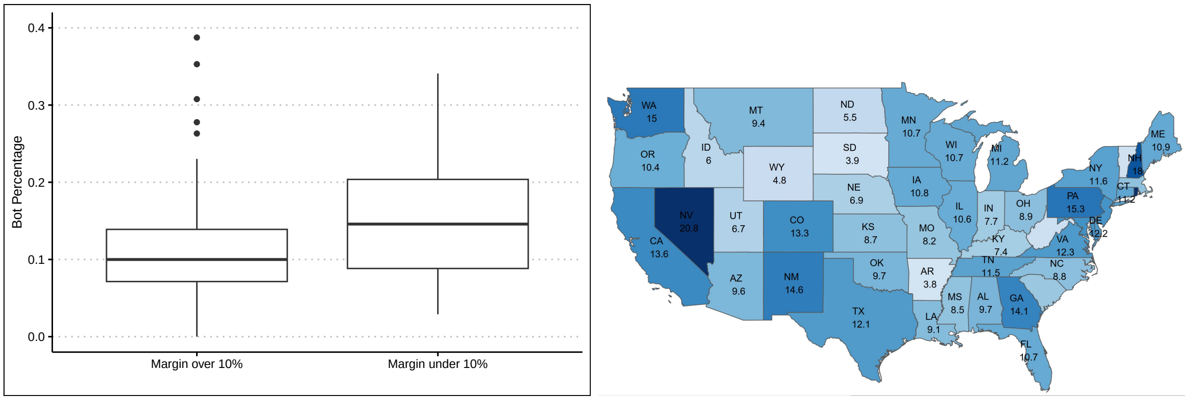
I theorize that politicians may exploit herding effects through artificially inflated social media popularity in close races or when targeting undecided voters. This strategy can be particularly effective in primaries, swing states, or low-salience local elections. Bot-boosted popularity could sway undecided or disengaged voters who use social media engagement as a decision-making shortcut.

I web scrape Ballotpedia to parse all the congress members and senate candidates who ran in Nov 2022, the second place candidate, and vote shares of each of these candidates. I then regress the margin of victory calculated from these against bot percentage and find that in close races where the margin of victory is under 10 percent, there is about 4.5% higher bot activity than when the race is not close. The raw regression is attached in the supplemental material. The map in panel B of Figure 3 shows the average bot percentage calculated from House and Senate members' Twitter pages in each state. A bird's eye view of the map also suggests swing states such as Georgia, Nevada, and Pennsylvania have high bot engagement rates. There is solid evidence linking the closeness of a race and bot percentages.[8] This analysis aims to describe politicians using various metrics rather than disentangling the effects of close elections and swing voters.

[7] The results are slightly larger but substantively the same when using ideology scores from GovTrack.us as opposed to DW-Nominate scores

[8] Interestingly, NH is not a swing state but saw a very close election between Maggie Hassan and Don Bolduc

Figure 3: Close races and swing states



Note: Some state labels have been omitted to maintain visual clarity. Please refer to the supplemental index for the entire table.

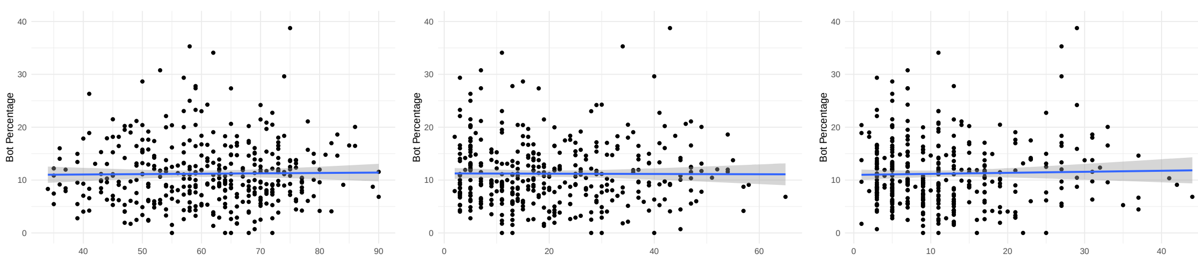
Age, number of years in office, number of years since first held office

Years in office data is webscraped from Ballotpedia, using the value under the candidate profile. However, many candidates are career politicians who may be running in new districts or different races. To account for this, I created a measure called ”years in political career,” which calculates when a candidate first held any public office until 2024.

For example, Chuck Schumer has spent 25 years in his current position, but his political career began in 1975 when he started his tenure in the New York State Assembly (1975-1980). Thus, his years in public service would be calculated as 49. This variable, ”years in public service,” is plotted against bot percentage in Figure 4, panel 3.

My analysis suggests that candidate age and experience do not impact bot engagement. Younger candidates or those with less experience are not any more likely to have high bot engagement rates. Figure 4 demonstrates that neither age, years in office, nor years since first assuming public office affects bot engagement.

Figure 4: Age and experience



Note: There is no relationship between candidate age and experience and bot percentage

Money in Politics and bot percentage

The following analysis uses Database on Ideology, Money in Politics, and Elections (DIME) data by Bonica (2015). I am using donation records from 2019-2022 to create estimates of the total amounts raised by each candidate from the DIME contributor database. The key variables I create for each candidate are the number of individual donors who donated to the campaign, the number of corporate and trade organizations donating to the candidate, and the dollar amounts raised by each candidate from corporate and individual sources.

The simple regression results[9] suggest that neither the amount raised by individual or corporate sources nor the number of donors from corporate or individual sources is very predictive of the bot engagement rates. However, the ratio of the counts of donors, i.e., the number of corporate donors divided by the number of individual donors, is significantly predictive of the bot percentage. This ratio is low for a candidate with a large denominator, i.e., a significantly large number of actual humans or individuals who have donated to this candidate's campaign. On the other hand, the proportion is closer to one for (or greater than one for some candidates) those who have barely any individual donations but a significant number of corporate donors.

The compelling aspect of this analysis is that the ratio of money from individual or corporate sources did not matter as much as the ratio of the number of donors of each type. As the ratio increases by 1, the bot percentage increases by almost 9%, which is a significant effect of approximately a 1.3 standard deviation increase in bot percentage. I interpret this finding to suggest that candidates with a lot of money but negligible grassroots support from actual voters (indicated by the lack of individual donations) have an incentive and the crucial financial resources to purchase inauthentic accounts and astroturf to level the playing field.

Measurements of online popularity

The **GDELT Online News Summary** is a part of the much larger GDELT project, which provides real-time monitoring of global news media to map events, actors, and trends. GDELT project scrapes approximately 150,000 to 200,000 news websites daily. These sources range from news sites, blogs, and discussion forums in various languages. Moreover, GDELT creates a whole host of additional metadata such as entity recognition (to label actors, locations, etc., associated with a news article), sentiment analysis, category of news, etc, as part of its own analysis of the dataset available to researchers. In this project, however, I string search a politician's first and last name in the raw news text data using the GDELT API. An example snapshot of the raw dataset for Congress member Matt Gaetz is attached in the appendix I. Similar files are created for every politician. The raw output JSON files have the date, the URL, and the title of these articles.

The final panel dataset tracks politician popularity over time by aggregating the number of articles mentioning each politician in either the text or title. This daily count serves as a proxy for

[9] Full regression results attached in the appendix

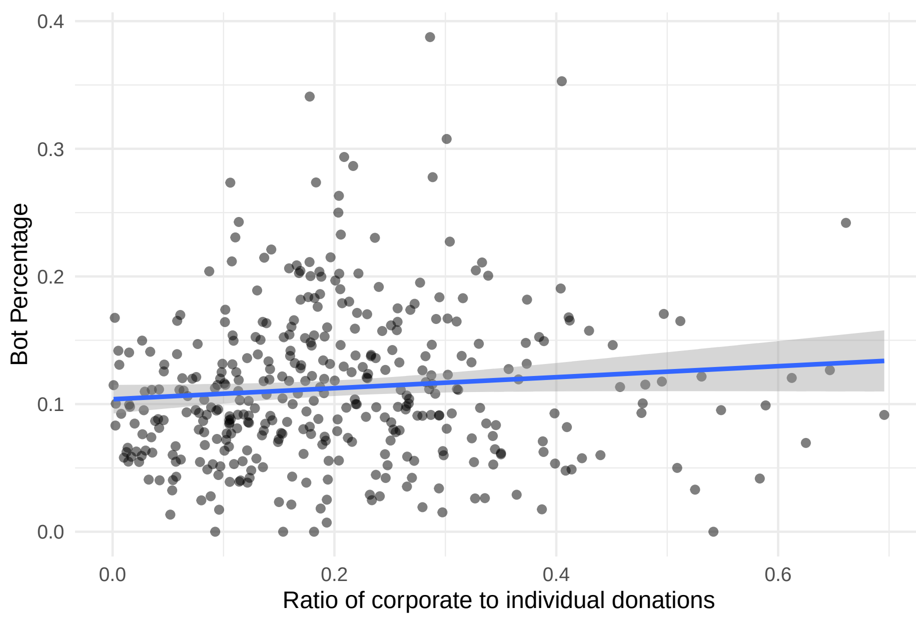


Figure 5: Higher ratios suggest fewer individual donations and more corporate donations which correlates with higher bot engagement rates

politician popularity, quantifying the frequency of media coverage for each political figure. This approach provides a measurable indicator of a politician's media presence and, by extension, their public visibility on a day-to-day basis

Appendix H presents summary statistics of the raw data, showing politicians are mentioned in approximately five sources daily on average, with considerable variation. For analysis, raw mention counts are normalized by dividing by the total number of articles GDLET scraped that day and multiplying by 100,000. This data is then merged with information on politicians' bot usage. Politicians are categorized based on their bot usage: high bot users (above median bot usage), low bot users (below median bot usage), and negligible Twitter users (insufficient data for reliable measurement, assumed to have minimal bot assistance due to low engagement). Figure 6 displays a scatterplot of logged news mentions counts against time for all politician categories. The analysis in the next sections compares high and low bot users, with an alternative comparison grouping low bot and negligible Twitter users together against high bot users. Both specifications yield similar findings.

American Television Global Knowledge Graph dataset is also part of the GDELT project[10] and this data source has the complete raw closed captioning of each monitored news show, covering every word spoken on major national and local news channels comprising more than two million hours of television news totaling more than 5.7 billion words from 163 distinct stations spanning

[10]The GDELT American Television backend dataset was aggregated by scraping data from Internet Archive's Television News Archive

July 2009 to present. For this paper, I analyzed the data for CNN, MSNBC, FOX News, FOX Business, Bloomberg, CSPAN, CSPAN2, and CSPAN3. I query the GDELT TV API and string search[11] the first and last name of every politician in my dataset to get a quantifiable proxy of how much they are being discussed on any channel. Politicians do not necessarily have to appear on T.V. If their names are mentioned in any context, it is counted. The objective is to understand if social media manipulation tactics show unexpected spillover effects on conventional T.V. news coverage. The output is a JSON file with the number of times the politician's name appeared per day per channel between Jan 2022 and Feb 2024.

Appendix table K presents the summary statistics of the raw daily name mentions per channel for 533 politicians over 776 days from January 2022 to February 2024. On average, politicians see about 0.17-0.27 name mentions per day (with a median of 0) on mainstream news channels such as CNN, FOXNEWS, and MSNBC. This data is sparsely populated because a small number of political figures receive extensive screen time on national television, while the majority are mentioned infrequently.

EMPIRICAL STRATEGY

My empirical strategy leverages a Twitter policy change that primarily affected politicians with high bot engagement on their pages. The bots examined in this study relied on the Twitter API to function, and the policy change disrupted their operations. First, I present results from an event study where I've binarized bot percentages and binned politicians into high and low bot users and compared their follower counts before and after the policy change. Next, I compare high and low bot engagement politicians' news mentions, which are presented in two ways: a difference-in-differences analysis using a continuous bot percentage measure and an event study using a binarized bot percentage measure. Additionally, I discuss the marginal effects of bot percentage on news mentions.

First Stage Results: Twitter API shutdown

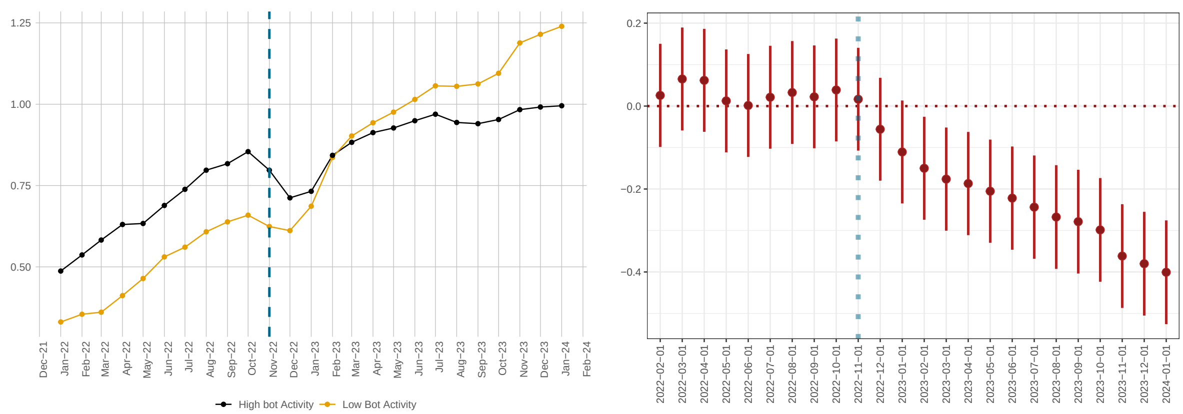
Elon Musk's acquisition of Twitter led to a period of significant upheaval for the platform. Between December 2022 and January 2023, the Twitter API experienced massive global outages and disruptions. The API disruptions notably targeted third-party applications like Twitterrific, Fenix, and Talon on both Android and iOS platforms, while Tweetdeck and complete clients were unaffected. Changes to the platform were confirmed by Elon Musk, who tweeted on December 28 that "Significant backend server architecture changes" had been implemented. Although the official announcement to terminate free API access and introduce a paid subscription model came

[11] Some politicians who have common names, such as Bill Johnson, for example, could yield incorrect estimates and have been excluded in the analysis.

in February 2023, many applications and services had been breaking since earlier as the platform probably tested these changes before the official announcement.

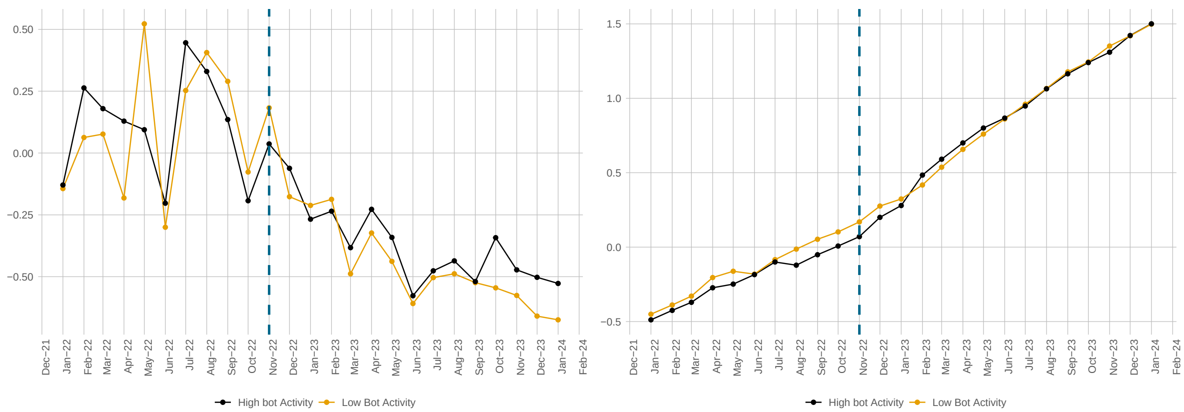
To understand the impact of the API shutdown, I assembled a dataset that tracks the total number of followers of a politician's account, which is averaged monthly over time. The average number of followers for politicians' accounts was 226,297 in Nov 2022, with a median of 35,649. There is significant variation with a minimum number of followers at 244 and a maximum at 12,642,266. In Figure 2, the left panel, the follower numbers are Z score normalized (subtracted by a mean number of followers and divided by the S.D. of every politician) to make visualization easier. Most politicians attempt to grow their social media, and typically, the number of followers is expected to increase gradually compared to the baseline of Jan 2022.

Figure 6: Parallel trends in the normalized follower counts before the intervention



Note: The politicians above the median bot usage are labeled as those having high bot activity. The graph shows how both types of politicians had a parallel trend in the number of normalized followers, which changed dramatically post-Twitter API infrastructure changes. The blue dotted line is the first reported month of the global outage. Although the policy was officially announced in Feb 2023, it had been breaking for months before that as it was being tested pre-implementation. Elon Musk took over Twitter management in Oct 2022, which caused some people to leave the platform. However, it is important to note that the trends were still parallel before the API changes started rolling out. The API shutdown was officially announced in Feb 2023.

Figure 7: Placebo checks to show no decline in engagement on Facebook and Instagram



Note: The politicians above the median bot usage are labeled as those having high bot activity. The graphs here demonstrate that the decline in the number of followers on Twitter was not observed on any other major social media platform. The first graph shows the number of (normalized) 'likes' a politician received on their Facebook post. The second-panel figure is more comparable conceptually to Figure 2 as it reports the number of normalized followers for a politician's Instagram account, which is very similar to the number of followers on Twitter. There is a steady, gradual increase, as expected.

The year preceding the API shutdown showed remarkably balanced parallel trends in follower numbers between accounts with high and low bot percentages. Worldwide outages began in November 2022 and continued to worsen, especially for third-party app users who do not contribute to Twitter's revenue. The bots using third-party software or the free APIs were naturally hit hard.

Analysis of follower trends in the 11 months leading up to the API disruption reveals that accounts with higher bot percentages maintained a larger follower base. However, in the wake of the API shutdown, this landscape changed significantly. The gap between high-bot and low-bot accounts not only closed but reversed, with this new trend continuing to grow. This shift suggests a significant impact of the API changes on the artificial inflation of follower counts, potentially exposing the extent to which bot activity had previously boosted certain accounts' popularity.

Next, the results of an event study analysis, which empirically examines the difference between high and low bot users each month, are summarized in Figure 2. The event study confirms my visual intuition that follower counts in 2022 pre-API shutdown showed parallel trends and post-shutdown showed diverging trends. Specifically, I am estimating the following non-parametric difference in the difference model with staggered treatments where the timing of the treatment assignment can vary across units. I estimate :

$$Y_{c,t} = \gamma_c + \phi_t + \sum_{n=-11}^{n=14} \beta_n * T_c * \phi_t^n + \epsilon_{c,t}$$

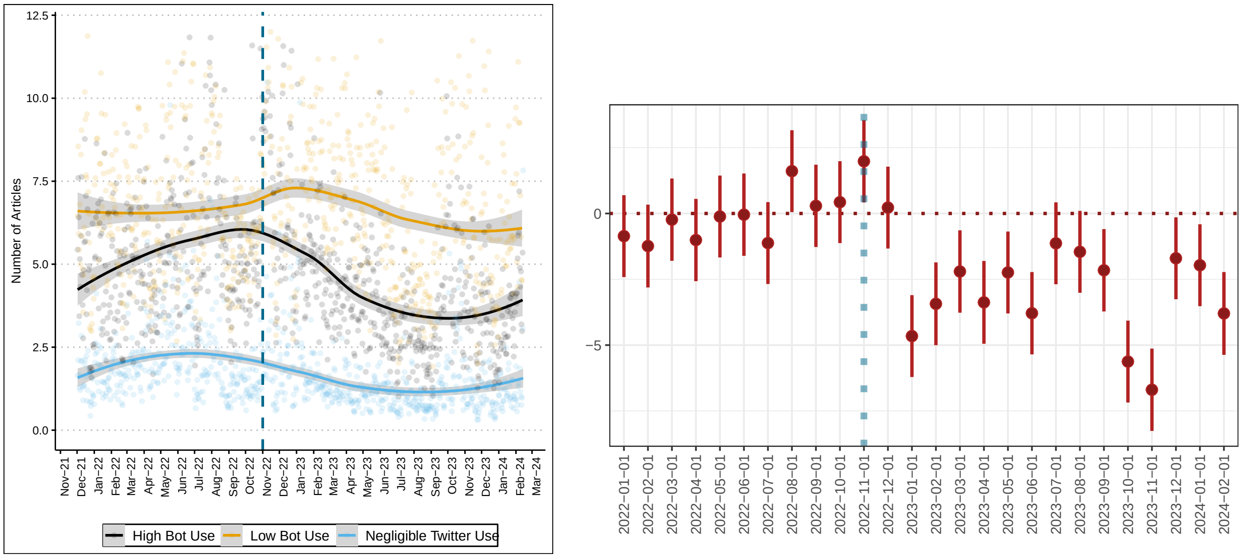
Here $Y_{c,t}$ is the outcome of interest, i.e., the normalized count of followers of politician c at the time (month) t. γ_c are the dummies that account for politician fixed effects, and ϕ_t are the dummies that account for time (month) fixed effects. $T_c = 1$ denotes treatment or high bot politician and $T_c = 0$ denotes low bot politician. The main coefficient of interest shown in Figure 2 is the interaction effect, which tracks the average difference in the outcome variable Y (normalized follower counts) between politicians in treatment and control groups that were n time periods away from the shock. A similar analysis is also done for the Facebook likes data and the Instagram followers data and is attached in the appendix section G

The analysis presented above supports two key claims. Firstly, it demonstrates that a significant proportion of Twitter bots available for purchase online or through digital marketing agencies in 2022 were designed to operate using the public API. The shutdown of this API resulted in a substantial decrease in followers, specifically for politicians identified as high bot users pre-shutdown. Second, while individual account-level analysis through the Botometer API may contain errors or noisy measurements, aggregating the data at the politician level allows for a clear separation between high and low bot-using politicians. Thus, the above results and distinct differences in follower counts following the API shutdown lend credibility to my methodology.

Second Stage Results (Part I): Impact on popularity in online news articles

In the preceding section, I demonstrated significant exogenous platform shocks affecting the Twitter bot population. Following this logic, politicians who relied on bots to artificially inflate their popularity should experience declining interaction and virality on Twitter. I argue that this initial impact leads to a second-stage effect: these politicians would likely see a reduction in their coverage by newsrooms. This could be due to reduced algorithmic boosts, a perception of decreased popularity among voters, or diminished visibility to journalists who monitor social media for trending topics. My central hypothesis is that high bot-using politicians, following the API shutdown, should experience a measurable decline in their coverage across both traditional broadcast television news and online news articles.

Figure 8: Global News Articles



Note A: The graph shows the raw data of the logged numbers of online news articles written wherein a politician is featured. Politicians are broken up into high and low bot users if they fall above or below the median bot usage. The third category in blue is negligible Twitter users for whom I could not create reliable estimates because they are inactive on the platform. Logically, such users can be assumed to be low-bot users. The results presented consider both such cases where negligible Twitter users are pooled in with the low bot users and compared against the high bot users.

Note B: The graph shows the results of the event study analysis. The specification is as before $Y_{c,t} = \gamma_c + \phi_t + \sum_{n=-11}^{n=14} \beta_n * T_c * \phi_t^n + \epsilon_{c,t}$. Here $Y_{c,t}$ is the outcome of interest, i.e., the total number of articles per 100,000 monitored articles where politician c at the time (month) t was mentioned. γ_c are the dummies that account for politician fixed effects, and ϕ_t are the dummies that account for time (month) fixed effects. $T_c = 1$ denotes treatment or high bot politician and $T_c = 0$ denotes low bot politician. The main coefficient of interest shown in the figure above is the interaction effect, which tracks the average difference in the outcome variable Y between politicians in treatment and control groups from the n^{th} time periods. The results here pool the low Twitter users with the low bot users. The raw means per month for each group are also attached in the appendix.

Table 2

	Dependent variable:Average Daily Articles		
	Base(Covariates)	*Politician FE*	
	(1)	(2)	(3)
Bot Percentage	0.650		
	(0.771)		
Post Shutdown	1.259***	1.259***	0.196**
	(0.172)	(0.158)	(0.091)
Bot Percentage:Post Shutdown	−14.023***	−14.023***	−8.733***
	(1.040)	(0.955)	(0.649)
Observations	306,324	306,324	428,532
N	382	382	533
Adjusted R^2	0.117	0.255	0.255
Residual Std. Error	26.855 (df = 306318)	24.664 (df = 305941)	21.238 (df = 427997)
Note:			*p<0.1; **p<0.05; ***p<0.01

The first column presents baseline results of the diff-in-diff without any politician-fixed effects. Columns 2 and 3 represent the full specification $Y_{c,t} = T + Bot_c * T + \phi_c + \gamma_{MMYY} + \epsilon_{c,t}$ with the politician fixed effect. Column 2 shows the comparison of politicians for whom I have accurate measurements, i.e., 382. The politicians who are negligible Twitter users are included in column 3 and pooled in by encoding their bot percentage as 0. The results remain robust.

The regression results in Table 2 presents findings from my Difference-in-Differences analysis[12], using the equation:

$$Y_{c,t} = T + Bot_c * T + \phi_c + \gamma_{MMYY} + \epsilon_{c,t}$$

where $Y_{c,t}$ is the number of politician name c mentions in digital news per day t; $T = 1$ is a binary variable that denotes if the observation is after 30 Nov 2022 and 0 otherwise; Bot_c is the bot percentage active on politicians c account. I also account for overarching time trends by month-year fixed effect. In Columns 2 and 3, I include politician fixed effect to focus on within-politician variation in online article name mentions over time. These fixed effects also help me account for unobserved individual heterogeneity and time-invariant characteristics of the candidates that may affect the online article name mentions but are not directly observable or measurable to reduce omitted variable bias. Column 3 includes all the politicians who are negligible Twitter users, pooled in by encoding their bot percentage as 0. The results remain robust and form the lower-bound estimate of my results, further validating my findings.

I observed a drop in articles mentioning high bot-using politicians post-Nov 2022. The effect sizes are equivalent to a decrease of 0.57 standard deviations (Column 2) and 0.35 standard

[12]The results of a more detailed event study analysis for this same dataset are summarized in Figure 4 on the right to show monthly variation in online news coverage between high and low bot using politicians

deviations (Column 3). Comparing the low bot-using politicians (with a bot percentage of 6% or the upper limit of the first quantile), I report a decrease of 0.32 articles mentioning the politician in the post-API shutdown phase. In contrast, high bot-using politicians (with a bot percentage of 20% or the lower limit of the third quantile) would see a much more sizable decrease of 1.55 daily articles mentioning their names.

It's important to remember that these estimates represent the lower bound of potential treatment effects, as all politicians with negligible Twitter usage are classified as low bot users, with their bot percentage set to 0 in the dataset for the third column results. The observed decrease of 1.55 daily articles mentioning a candidate represents a significant blow to their visibility as these numbers can quickly add up. Over the 441 days post-API shutdown, a high bot-using politician likely experienced 682 fewer articles mentioning them (equivalent to 46.5 fewer monthly articles). To contextualize this impact, the average politician in the dataset is mentioned in about five articles daily across all sources, totaling approximately 150 articles per month. Thus, a reduction of 46 articles per month constitutes a substantial decline, which could arguably hurt their popularity or relevance in the short-lived memories of voters.

An event study analysis, an extension of the standard difference-in-differences approach described above, allows me to dynamically illustrate the API shutdown's sustained effect over multiple months. These results provide a more nuanced view of the impact, showing how the effect evolves over time rather than presenting a single average effect. The results from the event study are summarized in Figure 4, and the full set of results is available in the supplemental material.

Decline in coverage is coming from lower quality sources

The previous section had a massive corpus of about 3.4 million articles that mentioned politicians. Next, by leveraging resources from Lin et al. (2023), I quantify the quality of news domains. The authors in this study analyze the variations in ratings from different media organizations and fact-checkers and found a high correlation among six sets of expert ratings. They then create a comprehensive set of domain ratings using an ensemble "wisdom of experts" approach which they make available to researchers here.

The data source provides quality ratings for 11,520 news sources. By matching the website host, subdomain, and suffix, I could assign quality attributes to about 31% of my data. This matched subset represents a non-random sample of the full corpus. The following figure illustrates the distribution of quality scores in the matched dataset. While Lin et al. (2023) generally recommend considering sources with a quality rating above 0.7 as high quality, I conservatively chose 0.75 as the threshold to bin high and lower quality sources of news coverage.

Table 3: Top News Sources and their volume in the dataset

High Quality		Low Quality	
Source	Percentage	Source	Percentage
MSN	12.30%	Yahoo	8.46%
Politico	3.90%	Breitbart	3.11%
Washington Post	2.19%	Fox News	2.98%
Newsweek	1.62%	Washington Times	2.45%
ABC NEws	1.42%	Washington Examiner	2.06%
AP News	1.31%	The Epoch Times	1.94%
New York Times	1.20%	Daily Mail	1.62%
NBC News	1.18%	Free republic	1.14%

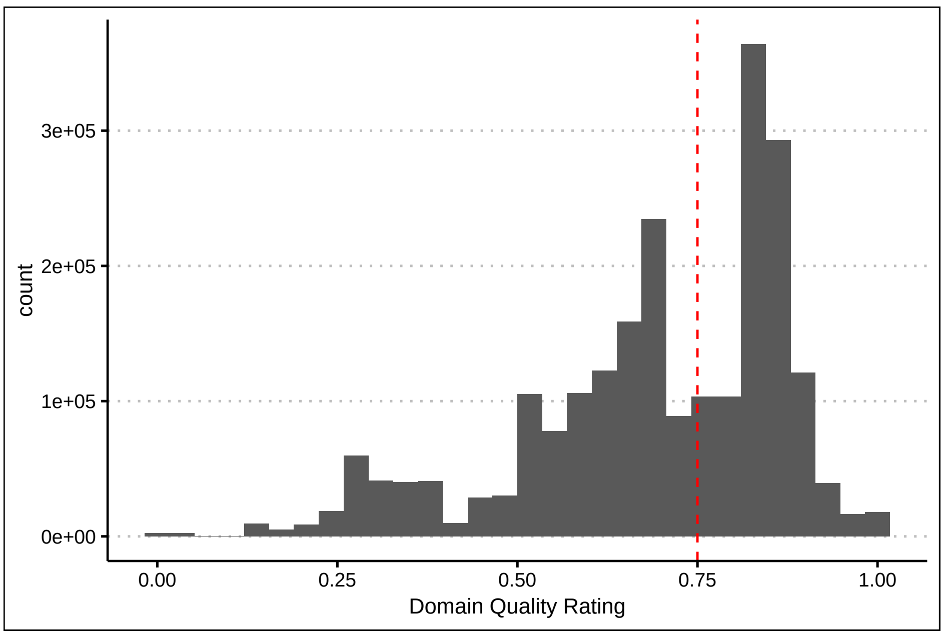


Figure 9: Distribution of domain quality of the news sources

The regression analysis setup presented here mirrors the previous difference-in-differences approach, with the key distinction being that the data is now subset to include only high or low-quality sources. The results below demonstrate a decline in coverage of high bot engagement politicians in the post-shutdown year from both types of sources. However, this decline is approximately three times larger when focusing solely on lower-quality domains as compared to high quality domains. This aligns with the intuition that lower-quality sources may rely more heavily on social media popularity when making editorial decisions.

	Dependent variable: Number of Daily Articles	
	High Quality	Low Quality
Post Shutdown	0.274^{***}	1.167^{***}
	(0.061)	(0.072)
Bot Percentage:Post Shutdown	-1.351^{***}	-3.869^{***}
	(0.495)	(0.583)
Observations	145,621	154,064
Adjusted R^2	0.143	0.144
Residual Std. Error	5.514 (df = 145242)	6.662 (df = 153685)
		Note: *p<0.1; **p<0.05; ***p<0.01

Second Stage Results (Part II): Impact on popularity in TV news coverage

The figure below descriptively shows the average values of politicians' name counts per month grouped by high and low bot-using politicians (above or below the median) and negligible Twitter-using politicians from Oct 2021 to Feb 2024. The dotted line shows the Twitter API disruption from Nov 2022.

Next, the regression results in Table 3 in the first panel present the results of 382 politicians and compare the high bot users against the low bot users. The empirical specification is

$$Y_{c,t,K} = T + Bot_c * T + \phi_c + \gamma_{MMYY} + \epsilon_{c,t,K}$$

where $Y_{c,t,K}$ is the number of politician name c mentions on channel K on day t; $T = 1$ denotes if the observation is after 30 Nov 2022 and 0 otherwise; Bot_c is the bot percentage active on politicians c account. I account for overarching time trends by month-year fixed effect. I include politician fixed effect to focus on within-politician variation in T.V. name mentions over time. These fixed effects also help me account for unobserved individual heterogeneity and time-invariant characteristics of the candidates that may affect the T.V. name mentions but are not directly observable or measurable to reduce omitted variable bias. The second panel estimates the same equation as above but includes all politicians who are negligible Twitter users to form the lower bound estimate of my results. Importantly, these results remain robust, further validating my findings.

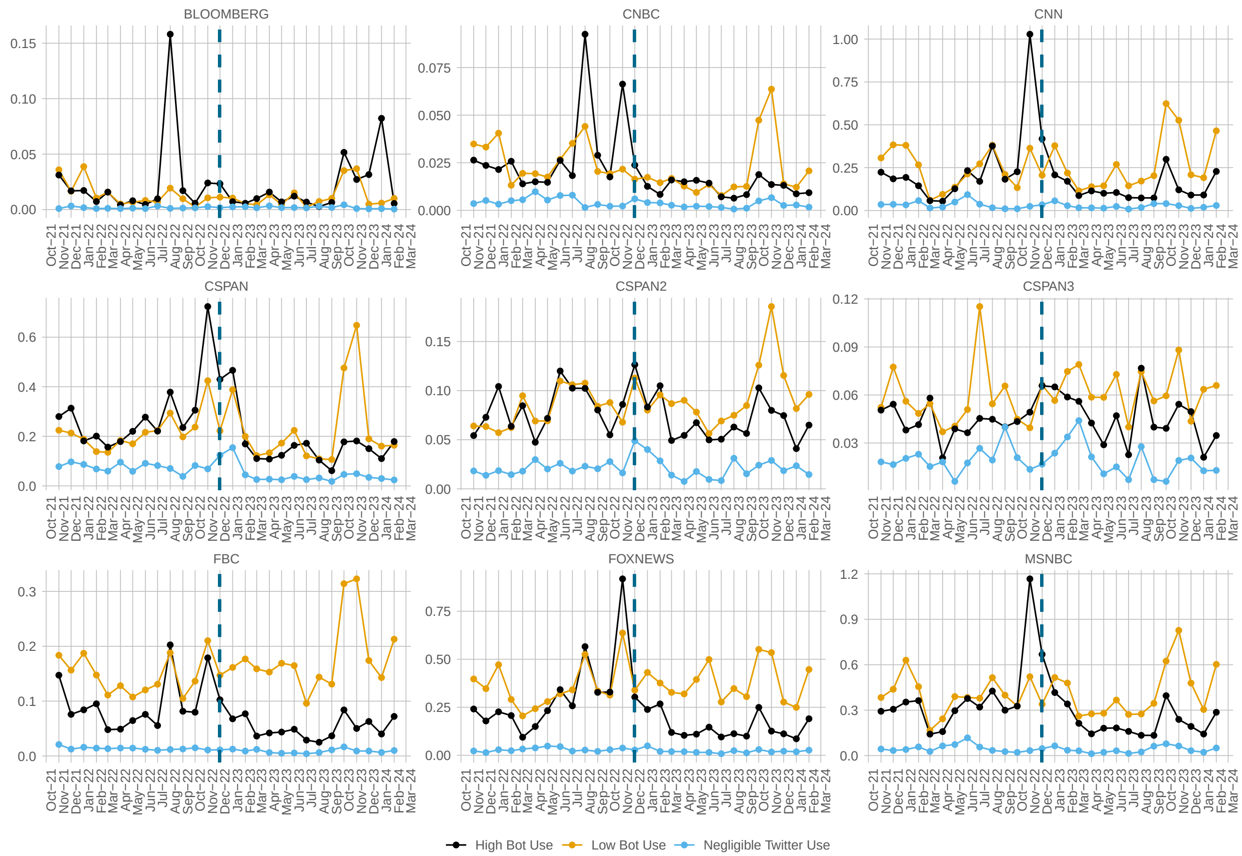

BLOOMBERG
CNBC
CNN
CSPAN
CSPAN2
CSPAN3
FBC
FOXNEWS
MSNBC
High Bot Use
Low Bot Use
Negligible Twitter Use

Table 4

	Dependent variable: Counts of name mentions on TV channel								
	Business			Politics				News	
	CNBC	FBC	BLOOMBERG	CSPAN	CSPAN2	CSPAN3	FOXNEWS	MSNBC	CNN
Post	0.007	0.052^{***}	0.007	0.130^{***}	0.047^{***}	-0.009	0.116^{**}	0.055	0.149^{***}
	(0.008)	(0.018)	(0.014)	(0.042)	(0.017)	(0.014)	(0.046)	(0.062)	(0.053)
Bot%×Post	-0.047^{***}	-0.319^{***}	-0.027	-0.544^{***}	-0.104^{***}	-0.024	-0.839^{***}	-0.847^{***}	-0.732^{***}
	(0.017)	(0.037)	(0.029)	(0.085)	(0.035)	(0.029)	(0.094)	(0.127)	(0.109)
N	302,133	302,133	302,133	302,133	302,133	302,133	302,133	302,133	302,133
Adj R^2	0.041	0.138	0.037	0.104	0.121	0.078	0.169	0.103	0.069
	Pooling Negligible Twitter Users:Lower Bound Results								
Post	0.009	0.025^{**}	0.005	0.108^{***}	0.045^{***}	-0.008	0.073^{**}	0.020	0.086^{**}
	(0.006)	(0.013)	(0.010)	(0.030)	(0.012)	(0.011)	(0.033)	(0.044)	(0.038)
Bot%×Post	-0.057^{***}	-0.191^{***}	-0.024	-0.390^{***}	-0.073^{***}	0.0004	-0.703^{***}	-0.688^{***}	-0.545^{***}
	(0.011)	(0.025)	(0.019)	(0.058)	(0.024)	(0.021)	(0.063)	(0.086)	(0.073)
N	421,083	421,083	421,083	421,083	421,083	421,083	421,083	421,083	421,083
Adjusted R^2	0.041	0.139	0.037	0.102	0.116	0.073	0.171	0.104	0.069

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

The first panel presents results of 382 politicians for whome I have resonable bot measurements. The emperical specification is $Y_{c,t,K} = T + Bot_c * T +_c + \gamma_{MMYY} + \epsilon_{c,t,K}$ where $Y_{c,t}$ is the number of politician name c mentions on channel K per day t; $T = 1$ denotes if the observation is after 30 Nov 2022 and 0 otherwise; Bot_c is the bot percentage active on politicians c account. I also control for relevant covariates, such as follower numbers and the number of tweets from the account, month-year fixed effect and politician fixed effect.

The second panel estimates the same equation as above but includes all the politicians who are negligible Twitter users to form the lower bound estimate of my results. These politicians are pooled in by encoding their bot percentage as 0. The results remain robust.

When looking at business news channels, I observe a decrease in name mentions post-Nov 2022 for politicians on CNBC and Fox Business among high bot users. The effect sizes are equivalent to a decrease of 0.15 standard deviations and 0.21 standard deviations for CNBC and Fox Business, respectively. To put these numbers in perspective, politicians who were low bot users at 6% active bots (i.e., the upper limit of the first quartile) would have seen an increase of 0.0133 daily name mentions on Fox Business in the post-API shutdown period. In contrast, politicians who were high bot users at 20% active bots (i.e., the lower limit of the third quartile) would have seen 0.013 fewer daily name mentions between Dec 2022 and Feb 2024.

Similarly, on CSPAN and CSPAN2, I observed a decrease in name mentions, equivalent to 0.19 standard deviations and 0.08 standard deviations. Thus, low bot-using politicians (say with a bot percentage of 6% or the upper limit of the first quantile) would observe an increase of 0.07 in name mentions, whereas high bot-using politicians (say with a bot percentage of 20% or the lower limit of the third quantile) would only observe a 0.02 increase in daily name mentions on CSPAN in the post API shutdown period.

Lastly, the effects are most prominent on mainstream news channels, where all three national channels in my dataset report a substantial decrease in name mentions in the post-API shutdown phase (Dec 2022 - Feb 2024).On CNN, MSNBC, and Fox News, I observed a reduction equivalent to effect sizes of 0.21 standard deviations, 0.23 standard deviations, and 0.31 standard deviations, respectively. Comparing the low bot-using politicians (with a bot percentage of 6% or the upper limit of the first quantile), I report an increase of 0.02 name mentions post API shutdown whereas, high bot-using politicians (with a bot percentage of 20% or the lower limit of the third quantile) would observe a decrease of 0.07 in daily name mentions on Fox News. Similarly, for CNN, low-bot-using politicians would see an increase of 0.04 daily name mentions, and high-bot politicians would observe a decrease of 0.028 daily name mentions.

It is important to note that these estimates represent the lower bound of potential treatment effects, as all politicians with negligible Twitter usage are classified as low bot users, with their bot percentage set to 0 in the dataset for the second-panel results. The observed decrease of 0.07 daily name mentions on a popular news channel such as FOXNEWS signifies a considerable impact on a candidate's visibility, as these numbers accumulate rapidly. To put these numbers in perspective, over the 441 days post-API shutdown, high bot-using politicians who probably enjoyed recognition on Fox News previously experienced 31 fewer name mentions in the subsequent 1.2 years (equivalent to about two fewer mentions per month) on this channel. Similarly, on CNN, high bot-using politicians saw approximately 13 fewer name mentions in the 1.2 years post-API shutdown, or about 0.8 fewer mentions per month on this channel. These reductions in media exposure across major news channels highlight the significant and lasting impact of the API shutdown on the visibility of politicians who heavily relied on bot boosting to maintain their media presence.

DISCUSSION

I present evidence in B that political elites without significant grassroots support and popularity (as measured by the number of individual donors to their campaign from the FEC data) and a lot of corporate funding (as indicated by the number of PACs or trade associations that donate to his/her campaign) have high rates of bot usage on their platforms. I interpret this finding to suggest that candidates with a lot of money but negligible grassroots support from actual voters have an incentive and the crucial financial resources to purchase inauthentic accounts and astroturf to level the playing field.

Normatively, bot usage significantly impacts American democratic institutions. Bots operate in a regulatory vacuum and do not need the consent of politicians to be active on their pages. Buying and using bots for campaigning and boosting popularity falls far outside FEC regulations. The ease with which anyone can purchase and utilize bot support for political campaigns, candidates, or petitions creates a landscape vulnerable to manipulation. There are attempts to regulate platforms, but solutions are complex, and different stakeholders have misaligned preferences.

While social media platforms have valid reasons[13] for being cautious about removing too many bot accounts, they have initiated efforts to curtail the impact of bots on their platforms. Low-effort Russian botnets, which could operate freely earlier, are banned quickly. Extensive bot networks originating from countries such as Russia, China, Venezuela, and Saudi Arabia have been identified and dismantled expeditiously. However, despite significant progress made by social media companies, they continue to grapple with challenges, often struggling to differentiate between automated bots and genuine users. Both companies and disrupters continue to evolve. The companies develop new detection systems, and bot programmers alter their behavior accordingly. For example, bot programmers are supposedly using AI-generated profile images instead of actual photographs or stock images, which are more easily traceable. D shows some stereotypical bot accounts and the red flag, which should make us suspicious of these accounts being real human users.

The United States currently lacks comprehensive federal laws for regulating social media bots. Creating strong laws for bots is challenging due to various legal and practical constraints. One major constraint is that any new law must respect the standards set by the First Amendment, which covers all regulations related to government-restricted speech. The constitutional protection given to bot speech is based on the idea that both people expressing their First Amendment rights and those receiving bot-generated messages have a right to share information under the Constitution. Even though this constitutional context does not prevent Congress from making rules, it highlights the complex nature of creating such regulations.

Some attempts have been made to regulate bots across various platforms, with one notable example being Sen. Dianne Feinstein's Bot Disclosure and Accountability Act. Introduced in 2018 and renewed in 2019, this bill aims to curtail bot usage by political entities and mandates clear identification of bots in social media posts. However, progress on this legislation has stalled in committee. Complicating matters further is Section 230 of the Communications Decency Act, which provides social media platforms immunity from legal liabilities arising from third-party content. This provision has been reinforced by the Trump administration's executive order addressing 'selective censorship,' potentially encouraging social media companies to leave up content they would otherwise consider for removal.

During his presidential campaign, Joe Biden proposed a complete repeal of Section 230, although this initiative has since lost traction. The potential consequences of such a repeal, if enacted without a suitable replacement, would be far-reaching. Social media platforms would suddenly find themselves legally responsible for all content posted on their sites. This shift in liability could prompt platforms to adopt extreme measures.

As regulations for bots are still in their early stages, the onus of managing bot activity primarily falls on social media platforms themselves. However, these companies often face conflicting incentives. On the one hand, bots can significantly boost engagement metrics, which in turn

[13] Removing too many bot accounts enmasse plummets engagement numbers which in turn reduces advertisement revenues and receives push back from investors

can positively impact stock values, making platforms reluctant to implement widespread bot removals. On the other hand, the aftermath of the 2016 election has ushered in a notable shift in platform behavior. Heightened public scrutiny and potential regulatory threats have created stronger incentives for platforms to actively combat bots, and both financial and reputational concerns drive these efforts.

CONCLUSION

My paper helps decipher an understudied actor (bot) on social media and its downstream effects on traditional media, which is assumed to be immune to such manipulations. Normatively, this research also has repercussions for American democratic institutions. My research aims to shift the focus from the conventional narrative that bots are solely used for spreading conspiracies or misinformation. The bots identified in this study may appear relatively benign as they are not directly propagating falsehoods. Instead, they contribute to creating an artificially inflated image of political candidates' popularity. This raises an interesting ethical question: Is this practice fundamentally different from paying people to attend a political rally, given that much of political engagement now occurs online?

Undoubtedly, malicious actors can and do use bots for nefarious purposes, as extensively documented in existing research. However, my contribution to this field lies in illuminating other aspects of inauthentic accounts that make them strategic tools in electoral campaigns. I would like to think that deliberate bots usage could influence early polling results or sway primary elections, contexts where more apparent determinants like partisanship have not yet come into play.

In this paper, I talk about three things primarily. First, I highlight the significance of social media popularity and how it can be artificially inflated to substitute for genuine grassroots support. Second, I demonstrate that Elon Musk's acquisition of Twitter and subsequent platform changes resulted in a significant exogenous shock to bot activity on the platform, causing a notable decrease in followers exclusively for politicians with high bot activity in the pre-shutdown period. My analysis of followers over time per politician and the first-stage results also validate my categorization of high and low bot users, creating what is, to my knowledge, the only dataset that systematically attempts such classification. Despite potential concerns about the account-level accuracy of the Botometer API used to identify inauthentic accounts, my aggregate analysis of hundreds of accounts retweeting each politician allows me to effectively separate the signal from noise on a per-politician basis.

Third, my key findings reveal a significant shift in online news coverage patterns for politicians post-November 2022. While high and low bot-using politicians had comparable coverage before this period, a diverging trend emerged afterward. After adjusting for seasonal variations post-Dec 2022, high bot-using politicians experienced a substantial decrease in name mentions across online news articles. Comparisons between high and low bot users consistently show significant effects, even when politicians with minimal Twitter presence are included as low bot users. The inclusion of negligible twitter users understandably attenuates my coefficients but forms the lower bound of potential effects and the results remain robust. Similar patterns are observed in broadcast TV coverage, with the most pronounced effects evident on mainstream news channels such as Fox News, CNN, and MSNBC.

In conclusion, bots and platforms are both evolving. Modern bots have sophisticated digital personas, complete with fake names, bios, and often AI-generated or stolen profile pictures. The type of bots used in the pro-Saudi content creation were low effort and were weeded out quickly by the platform's moderators. Unlike their predecessors, which were easily detected because of relentless propaganda, new-age bots carefully seed messages and mimic human behavior to avoid detection. These bots are deeply embedded in networks of actual humans and sometimes achieve status akin to political influencers. If the rise of deep-fakes and deep-fake detection technologies is to give us any hints, these digital personas will become more complex and more challenging to detect.

We are in a social media environment where bots are here to stay. They could be controlled by foreign actors, domestic political groups, campaign staff, or even political elites. In equilibrium, it will most likely be everybody. My objective with this research is to causally identify tangible real life effects beyond social media platforms and contribute to the debate on the seriousness of the issue.

References

Norah Abokhodair, Daisy Yoo, and David W. McDonald. Dissecting a Social Botnet: Growth, Content and Influence in Twitter. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 839–851, February 2015. doi: 10.1145/2675133.2675208. URL `http://arxiv.org/abs/1604.03627`. arXiv:1604.03627 [cs].

Nuha Albadi, Maram Kurdi, and Shivakant Mishra. Hateful People or Hateful Bots? Detection and Characterization of Bots Spreading Religious Hatred in Arabic Social Media. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–25, November 2019. ISSN 2573-0142. doi: 10.1145/3359163. URL `http://arxiv.org/abs/1908.00153`. arXiv:1908.00153 [cs].

Sitaram Asur, Bernardo A Huberman, Gabor Szabo, and Chunyan Wang. Trends in social media: Persistence and decay. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 5, pages 434–437, 2011.

Tim Bajarin. Russian-Ukraine War...The Most Broadcast War In History That Includes Physical And Cyber Warfare. URL `https://www.forbes.com/sites/timbajarin/2022/03/02/ukraine-russian-warthe-most-broadcast-war-in-history/`. Section: Consumer Tech.

Pablo Barberá, Andreu Casas, Jonathan Nagler, Patrick J Egan, Richard Bonneau, John T Jost, and Joshua A Tucker. Who leads? who follows? measuring issue attention and agenda setting by legislators and the mass public using social media data. *American Political Science Review*, 113(4): 883–901, 2019.

C Bell and A Coleman. Khashoggi: Bots feed saudi support after disappearance. *BBC News*, 2018.

Vera Bergengruen. Inside the Kremlin's Year of Ukraine Propaganda, February 2023. URL `https://time.com/6257372/russia-ukraine-war-disinformation/`.

Alessandro Bessi and Emilio Ferrara. Social bots distort the 2016 U.S. Presidential election online discussion. *First Monday*, November 2016. ISSN 1396-0466. doi: 10.5210/fm.v21i11.7090. URL `https://firstmonday.org/ojs/index.php/fm/article/view/7090`.

Adam Bonica. Database on ideology, money in politics, and elections (dime). *Harvard Dataverse*, 2, 2015.

Kellyton Brito and Paulo Jorge Leitão Adeodato. Measuring the performances of politicians on social media and the correlation with major latin american election results. *Government Information Quarterly*, 39(4):101745, 2022.

David A Broniatowski, Amelia M Jamison, SiHua Qi, Lulwah AlKulaib, Tao Chen, Adrian Benton, Sandra C Quinn, and Mark Dredze. Weaponized health communication: Twitter bots and russian trolls amplify the vaccine debate. *American journal of public health*, 108(10):1378–1384, 2018.

Lucy Butler. Beyond the individual: Examining how characteristics of social media information environments impact susceptibility to misinformation and the effectiveness of misinformation interventions. 2024.

Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web*, pages 675–684, 2011.

Georgiana Diana Ciocirdel and Mihai Varga. Election prediction based on wikipedia pageviews, 2016.

Devin Coldewey. Romney twitter account gets upsurge in fake followers, but from where. *NBC News*, 8, 2012.

David M Cook, Benjamin Waugh, Maldini Abdipanah, Omid Hashemi, and Shaquille Abdul Rahman. Twitter deception and influence: Issues of identity, slacktivism, and puppetry. *Journal of Information Warfare*, 13(1):58–71, 2014.

Brian de Silva and Ryan Compton. Prediction of foreign box office revenues based on wikipedia page activity. *arXiv preprint arXiv:1405.5924*, 2014.

Tomás Dodds, Claes de Vreese, Natali Helberger, Valeria Resendez, and Theresa Seipp. Popularity-driven Metrics: Audience Analytics and Shifting Opinion Power to Digital Platforms. *Journalism Studies*, 24(3):403–421, February 2023. ISSN 1461-670X. doi: 10.1080/1461670X.2023.2167104. URL https://doi.org/10.1080/1461670X.2023.2167104. Publisher: Routledge _eprint: https://doi.org/10.1080/1461670X.2023.2167104.

Erwan Dujeancourt and Marcel Garz. The effects of algorithmic content selection on user engagement with news on Twitter. *The Information Society*, 0(0):1–19, July 2023. ISSN 0197-2243. doi: 10.1080/01972243.2023.2230471. URL https://doi.org/10.1080/01972243.2023.2230471. Publisher: Routledge _eprint: https://doi.org/10.1080/01972243.2023.2230471.

Emilio Ferrara. Bots, Elections, and Social Media: A Brief Overview. In Kai Shu, Suhang Wang, Dongwon Lee, and Huan Liu, editors, *Disinformation, Misinformation, and Fake News in Social Media: Emerging Research Challenges and Opportunities*, Lecture Notes in Social Networks, pages 95–114. Springer International Publishing, Cham, 2020. ISBN 978-3-030-42699-6. doi: 10.1007/978-3-030-42699-6_6. URL https://doi.org/10.1007/978-3-030-42699-6_6.

Fabio Franch. (wisdom of the crowds) 2: 2010 uk election prediction with social media. *Journal of Information Technology & Politics*, 10(1):57–71, 2013.

Federica Giummolè, Salvatore Orlando, and Gabriele Tolomei. Trending topics on twitter improve the prediction of google hot queries. In *2013 International Conference on Social Computing*, pages 39–44. IEEE, 2013.

Raúl Gómez-Martínez, Carmen Orden-Cruz, and Juan Gabriel Martínez-Navalón. Wikipedia pageviews as investors' attention indicator for nasdaq. *Intelligent systems in accounting, finance and management*, 29(1):41–49, 2022.

Mitchell Amy Jurkowitz Mark Gottfried, Jeffrey and Jacob Liedke. Many journalists say social media helps at work, but most decry its impact on journalism. 2022.

Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. Fake news on twitter during the 2016 us presidential election. *Science*, 363(6425):374–378, 2019.

Jacob Groshek and Megan Clough Groshek. Agenda trending: Reciprocity and the predictive capacity of social network sites in intermedia agenda setting across issues over time. *Available at SSRN 2199144*, 2013.

Daniel Gruhl, Ramanathan Guha, Ravi Kumar, Jasmine Novak, and Andrew Tomkins. The predictive power of online chatter. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 78–87, 2005.

Jens Hainmueller, Jonathan Mummolo, and Yiqing Xu. How much should we trust estimates from multiplicative interaction models? simple tools to improve empirical practice. *Political Analysis*, 27(2):163–192, 2019.

Sounman Hong and Daniel Nadler. Which candidates do the public discuss online in an election campaign?: The use of social media by 2012 presidential candidates and its impact on candidate salience. *Government information quarterly*, 29(4):455–461, 2012.

Philip N Howard and Bence Kollanyi. Bots,# strongerin, and# brexit: Computational propaganda during the uk-eu referendum. *arXiv preprint arXiv:1606.06356*, 2016.

Stefano Maria Iacus, Giuseppe Porro, Luigi Curini, and Andrea Ceron. Every tweet counts? how sentiment analysis of social networks can improve our knowledge of citizens' policy preferences. an application to italy and france. Technical report, 2012.

David Ignatius. How a chilling saudi cyberwar ensnared jamal khashoggi. *Washingtonpost. com*, 2018.

Andreas Jungherr. Twitter use in election campaigns: A systematic literature review. *Journal of Information Technology & Politics*, 13(1):72–91, March 2016. ISSN 1933-1681. doi: 10.1080/19331681.2015.1132401. URL `https://www.proquest.com/docview/1784139607?parentSessionId=Nj%2BWTChCi%2FgB5TGV8gIvyySsFShCKArO0iDQEs6Uvi0%3D&pq-origsite=summon`. Num Pages: 72-91 Place: Abingdon, United Kingdom Publisher: Taylor & Francis Ltd.

Tobias R Keller and Katharina Kleinen-von Königslöw. Followers, spread the message! predicting the success of swiss politicians on facebook and twitter. *Social Media+ Society*, 4(1):2056305118765733, 2018.

Tobias R Keller and Ulrike Klinger. Social bots in election campaigns: Theoretical, empirical, and methodological implications. *Political Communication*, 36(1):171–189, 2019.

Bence Kollanyi, Philip N Howard, and Samuel C Woolley. Bots and automation over twitter during the first us presidential debate. *Comprop data memo*, 1:1–4, 2016.

Andrew Leber and Alexei Abrahams. A storm of tweets: Social media manipulation during the gulf crisis. *Review of Middle East Studies*, 53(2):241–258, 2019.

Hause Lin, Jana Lasser, Stephan Lewandowsky, Rocky Cole, Andrew Gully, David G Rand, and Gordon Pennycook. High level of correspondence across different news domain quality rating sets. *PNAS nexus*, 2(9):pgad286, 2023.

Juliane A Lischka and Marcel Garz. Clickbait news and algorithmic curation: A game theory framework of the relation between journalism, users, and platforms. *New Media & Society*, 25 (8):2073–2094, August 2023. ISSN 1461-4448. doi: 10.1177/14614448211027174. URL https://doi.org/10.1177/14614448211027174. Publisher: SAGE Publications.

Shannon C McGregor. "taking the temperature of the room" how political campaigns use social media to understand and represent public opinion. *Public Opinion Quarterly*, 84(S1):236–256, 2020.

Sharon Meraz. Using time series analysis to measure intermedia agenda-setting influence in traditional media and political blog networks. *Journalism & mass communication quarterly*, 88(1): 176–194, 2011.

Márton Mestyán, Taha Yasseri, and János Kertész. Early prediction of movie box office success based on wikipedia activity big data. *PloS one*, 8(8):e71226, 2013.

Panagiotis T. Metaxas and Eni Mustafaraj. Science and society. Social media and the elections. *Science (New York, N.Y.)*, 338(6106):472–473, October 2012. ISSN 1095-9203. doi: 10.1126/science.1230456.

Merja Myllylahti. Paying Attention to Attention: A Conceptual Framework for Studying News Reader Revenue Models Related to Platforms. *Digital Journalism*, 8(5):567–575, May 2020. ISSN 2167-0811, 2167-082X. doi: 10.1080/21670811.2019.1691926. URL https://www.tandfonline.com/doi/full/10.1080/21670811.2019.1691926.

Philip M. Napoli and Robyn Caplan. When Media Companies Insist They're Not Media Companies and Why It Matters for Communications Policy, March 2016. URL https://papers.ssrn.com/abstract=2750148.

Ben Nimmo. Fake news: Defining and defeating, Oct 2018. URL https://medium.com/dfrlab/fake-news-defining-and-defeating-43830a2ab0af.

John H Parmelee. The agenda-building function of political tweets. *New media & society*, 16(3): 434–450, 2014.

Nathaniel Persily and Joshua A Tucker. Social media and democracy: The state of the field, prospects for reform. 2020.

Sebastian Adrian Popa, Zoltán Fazekas, Daniela Braun, and Melanie-Marita Leidecker-Sandmann. Informing the public: How party communication builds opportunity structures. *Political Communication*, 37(3):329–349, 2020.

Jacob Ratkiewicz, Michael Conover, Mark Meiss, Bruno Gonçalves, Alessandro Flammini, and Filippo Menczer. Detecting and tracking political abuse in social media. In *Proceedings of the International AAAI Conference on Web and social media*, volume 5, pages 297–304, 2011.

Sippo Rossi, Matti Rossi, Bikesh Raj Upreti, and Yong Liu. Detecting political bots on twitter during the 2019 finnish parliamentary election. In *Annual Hawaii International Conference on System Sciences*, pages 2430–2439. Hawaii International Conference on System Sciences, 2020.

Hamza Salem and Fabian Stephany. Wikipedia: a challenger's best friend? utilizing information-seeking behaviour patterns to predict us congressional elections. *Information, Communication & Society*, 26(1):174–200, 2023.

Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. The spread of low-credibility content by social bots. *Nature Communications*, 9(1):4787, November 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-06930-7. URL http://www.nature.com/articles/s41467-018-06930-7. Number: 1 Publisher: Nature Publishing Group.

Bruno Castanho Silva and Sven-Oliver Proksch. Fake it 'til you make it: a natural experiment to identify european politicians' benefit from twitter bots. *American Political Science Review*, 115(1): 316–322, 2021.

Benjamin K Smith and Abel Gustafson. Using wikipedia to predict election outcomes: online behavior as a predictor of voting. *Public Opinion Quarterly*, 81(3):714–735, 2017.

Yan Su, Jun Hu, and Danielle Ka Lai Lee. Delineating the transnational network agenda-setting model of mainstream newspapers and twitter: a machine-learning approach. *Journalism Studies*, 21(15):2113–2134, 2020.

Joshua A Tucker, Andrew Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brendan Nyhan. Social media, political polarization, and political disinformation: A review of the scientific literature. *Political polarization, and political disinformation: a review of the scientific literature (March 19, 2018)*, 2018.

Andranik Tumasjan, Timm Sprenger, Philipp Sandner, and Isabell Welpe. Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Proceedings of the international AAAI conference on web and social media*, volume 4, pages 178–185, 2010.

Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. *Science*, 359 (6380):1146–1151, 2018.

David H Weaver, Lars Willnat, and G Cleveland Wilhoit. The american journalist in the digital age: Another look at us news people. *Journalism & Mass Communication Quarterly*, 96(1):101–130, 2019.

John P Wihbey. *The social fact: News and knowledge in a networked world*. Mit Press, 2019.

C Williams and G Gulati. Social networks in political campaigns: Facebook and congressional elections 2006. 2008.

Kai-Cheng Yang, Emilio Ferrara, and Filippo Menczer. Botometer 101: Social bot practicum for computational social scientists. *Journal of computational social science*, 5(2):1511–1528, 2022.

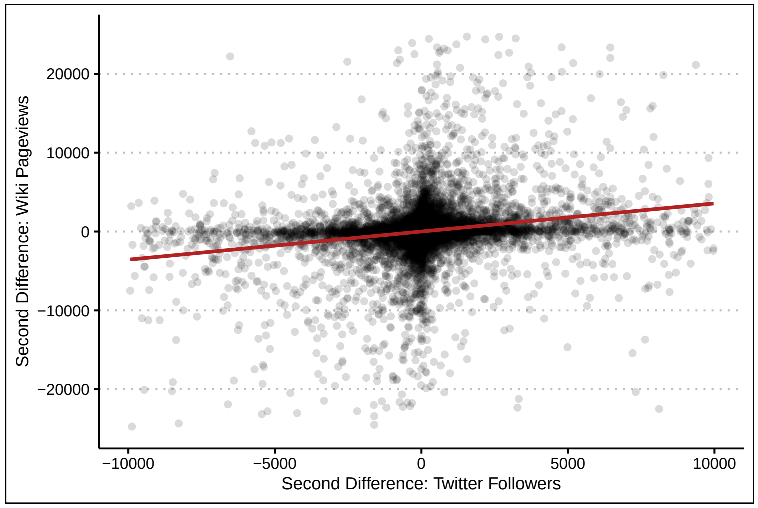
Taha Yasseri and Jonathan Bright. Wikipedia traffic data and electoral prediction: towards theoretically informed models. *EPJ Data Science*, 5:1–15, 2016.

A Appendix

Twitter is a potential trendsetter where political interests originate

Let us put aside the question of bot activity for a moment to understand the importance of social media numbers. This section illustrates the importance of Twitter's follower count in explaining and predicting additional web trends derived from Wikipedia. This prelimnary calculation demonstrate that interest originating in real-time on Twitter could anticipate interest in politicians in the weeks to come later on the web. I create two-time series for each politician, one which measures average weekly followers and another which is the average weekly Wikipedia pageviews for the same seven-day period.

This time series analysis of weekly Wiki pageviews allows me to understand the temporal correlation and the predictive power of the follower numbers at various time lags. I use the Granger causality test to understand if Twitter numbers predict Wikipedia pageviews or vice versa. Granger causality tests are standard in political science literature. (Groshek and Clough Groshek, 2013) use Granger relationships to understand if social networking sites shape media agendas on specific topics. (Su et al., 2020) use this technique to evaluate the relationship between newspaper coverage and the conversation around the extradition bill movement in Hong Kong on Twitter (and vice versa) and find that Twitter strongly influences newspaper coverage. (Meraz, 2011) also uses time series analysis to determine whether political blogs set the agenda for traditional news media



outlets and finds that traditional media could not set political blog agendas. However, blogs were able to set traditional media's online agenda. The figure below shows a simple scatterplot of the second-order differences in Twitter follower numbers and Wiki pageviews.

Finally, even though the statistical test is called a Granger causality test, I am not actually testing for causality in the traditional sense. Instead, the purpose of this exercise is to determine if the incremental forecasting value of Twitter follower numbers (and its lags) correlates with another Wikipedia pageviews within the model; i.e can the time series of Twitter followers be used to forecast Wiki pageviews (and vice versa).

The analysis below is simple; I begin with vector autoregressive (VAR) models. I use differencing (computing the differences between consecutive observations) to ensure that both time series are stationary. I also use the Akaike information criterion (AIC) to identify optimal lag length but report results in the table for multiple lags. The results below show that 37% of the politicians in my dataset have their Twitter following numbers granger causing their Wiki pageviews at one week time lag at 0.05 percent significance level. The opposite relation is much weaker, with only 11% of the politicians whose Wiki pageviews *granger cause* Twitter followers at the same significance level and lag. Even at two, three and four week lags, Twitter follower numbers consistently predict Wikipedia pageviews for a larger share of politicians than the other way around.

Order of Lags (Week)	Twitter → *Wiki*	Wiki → *Twitter*
1	0.37	0.11
2	0.28	0.11
3	0.32	0.11
4	0.35	0.16

B Politicians with significant financing and not enough grassroots support use bots

Table 5

	Dependent variable: Bot Percentage			
	(1)	(2)	(3)	(4)
$\frac{Money_{Corp}}{Money_{Ind}}$	-0.01			0.003
	(0.01)			(0.01)
$\frac{DonorCount_{Corp}}{DonorCount_{Ind}}$	0.06^{**}			0.09^{***}
	(0.03)			(0.03)
$log(Money_{Corp})$		-0.0003		-0.01
		(0.004)		(0.005)
$log(Money_{Ind})$		0.005		0.01^{*}
		(0.004)		(0.01)
$DonorCount_{Corp}$			0.0000	-0.0000
			(0.0000)	(0.0000)
$DonorCount_{Ind}$			0.00	0.0000
			(0.0000)	(0.0000)
Constant	0.11^{***}	0.05	0.11^{***}	0.02
	(0.01)	(0.04)	(0.01)	(0.05)
Observations	382	382	382	382
R^2	0.02	0.02	0.02	0.05
Adjusted R^2	0.01	0.004	0.01	0.02
Note:		*p<0.1; **p<0.05; ***p<0.01		

C Relationship between number of followers and bot percentage

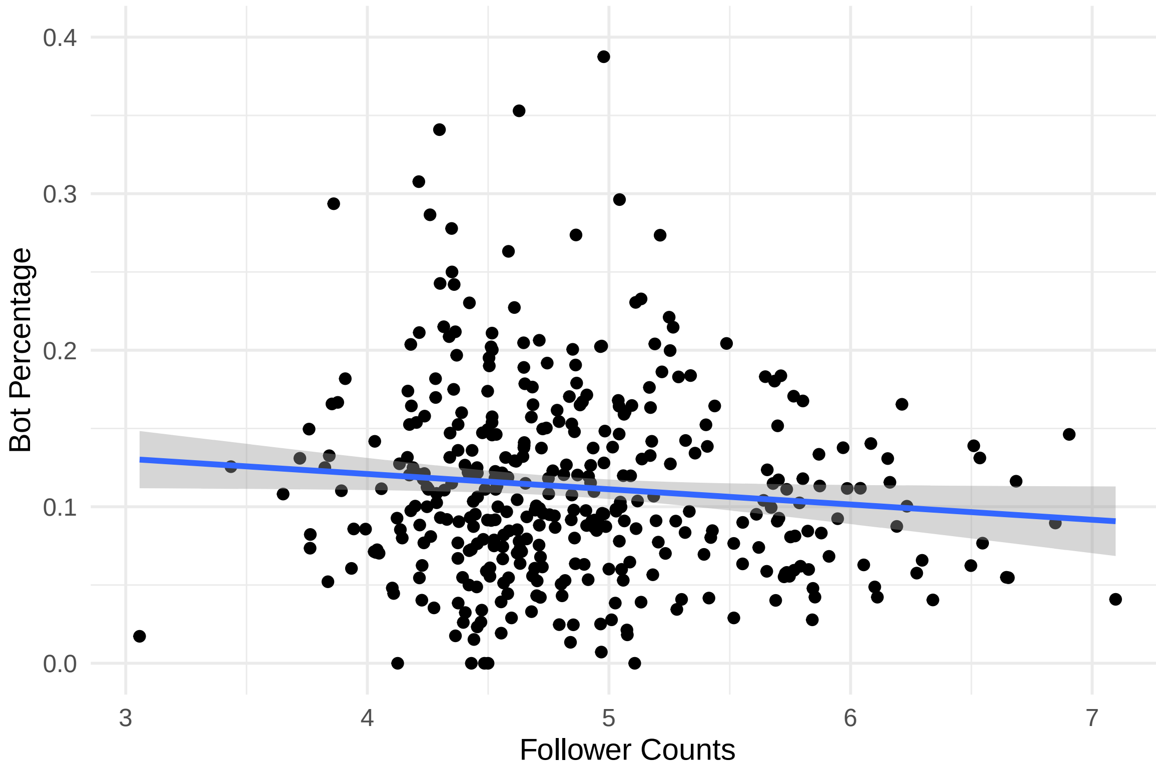


Figure 10: This *slight* negative statistically significant relationship suggests that politicians with very high follower counts in Nov 2022 did not rely on bot engagement or perhaps would have found no marginal benefit from increaed bot interaction. Follower counts presented above are in log_{10} for ease of interpretation

D Screenshots of potential bot accounts and associated red flags

The following account screenshots are typical bot accounts that a regular Twitter user might encounter. It is standard for bots not to have profile pictures or to use generic online pictures as their profile pictures, as seen in Bot 1, 4, and 5. Other times, the profile picture may be stolen from an actual account without the account holder's knowledge. A reverse image search shows that Bot 2's profile picture comes from this Instagram account belonging to Amber Mirza. Some stolen pictures could be slightly photoshopped to make reserve image searching difficult. For example, bot 3 below has a pair of obviously fake glasses photoshopped onto an image. The profile image on Bot 6 is a stock image from this source under the caption of 'portrait of happy woman with blowing hair' and is also the first image that shows up when searching for this description.

The next red flag to look out for is the number of posts the account makes. Bot 1 has a total of 467,3000 posts, meaning this account has tweeted about 150 times a day every day for the last eight years. Bot 4 has tweeted 382,700 times since Jan 2022, approximately a whopping 420 tweets daily. This is an abnormally large number of posts. For reference, the NPR Politics and New York Times Twitter channels, which have specialized teams of social media managers and were created in 2007, have about 111,000 posts and 545,000 posts, respectively. The point of a bot account is to create volume, so almost all of these messages are re-posts from other authors with next to no original tweets.

Other minor and subtle signs include the number of followers or following ratios. Most bots follow many different people but have few followers in return. Their profile names often have a long string of numbers following a name. They do not usually have a location and either have empty profile descriptions or very long emoji-filled descriptions, which avoid detection and suspicion by Twitter's internal checks. It is also essential to note that many low-effort bots on the platform can be used to bulk up someone's follower numbers or get easy 'likes'. These are easy to spot and are taken down just as fast. These bots usually have low follower and following counts and make few posts. The following section, which compares bots to humans, shows the bimodal nature of the followers, following, and tweets only seen in bots, not human accounts.

E Bots compared to human users

The following figures and tables look at the distribution of social media characteristics (such as the number of tweets accounts write, the number of accounts they follow, and the number of followers) of bots compared to regular human users. Statistics for authentic human accounts are shown in blue, compared to a bot account in red. An interesting pattern to note here is that bot accounts, on average, display a bimodal trend, whereas human account behaviors are more normally distributed. Furthermore, these graphs also suggest that bots remaining on the platform are not low-effort bots researchers might have encountered five years ago in the Jamal Khashoggi case mentioned earlier. Presumably, most social media platforms regularly purge low-effort bots. These inauthentic accounts are deeply embedded in human networks (on average, bots have 4231 followers with a median of 865) and much more active than real humans on the platform. Moreover, it serves as a preliminary validation that the accounts I have identified as inauthentic have remarkably different characteristics compared to human users and have been created much more recently.

((a)) Bot 1

((b)) Bot 2

((c)) Bot 3

((d)) Bot 4

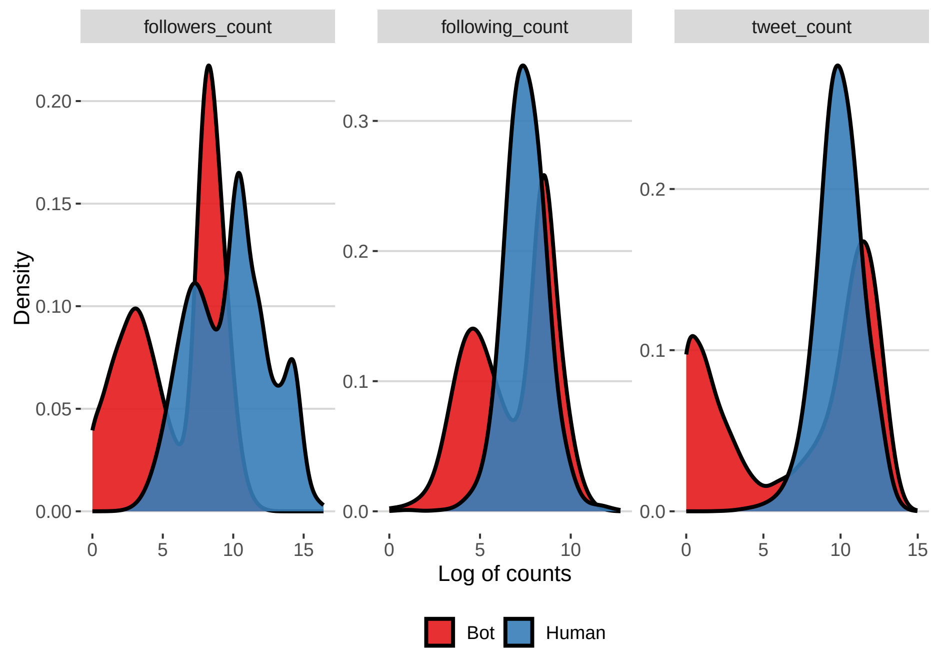
((e)) Bot 5

((f)) Bot 6

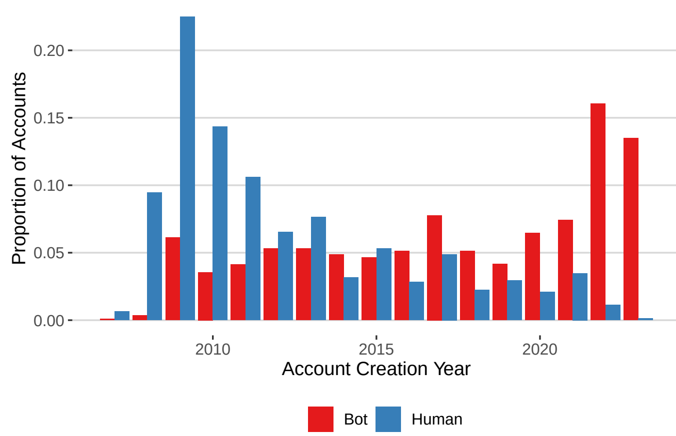
Figure 11: Potential Bot accounts and tell tale red flags

Figure 12: Distribution of the number of followers, following and total tweets by bots and humans

((a)) Followers, Following and Total Tweets



((b)) Year the account was created



Note: The following three figures show the count of logged numbers of followers, the logged number of following, and the logged count of total tweets since the account was created. This serves as a preliminary check that the accounts identified through the API behave very differently than regular human accounts.

Table 6: Summary Statistics of Bot vs Human accounts

	N	Mean	SD	Median	Min	Max
Bot Account						
Bot API Score	39371	4.54	0.24	4.5	4.2	5.0
Followers count	39371	4231.91	9217.24	865	0	219882
Following count	39371	4410.25	7965.77	1550	0	170600
Total Tweets	39371	73877.23	139451.65	10221	0	2186429
Human Account						
Bot API Score	103089	1.24	0.82	1.2	0.0	2.9
Followers count	103089	291162.1	881105.56	22725	1	13415758
Following count	103089	3959.02	12247.37	1747	0	341372
Total Tweets	103089	52507.10	93106.97	21552	0	1768819

Note: This is a comparison of bot and human accounts. It is important to note that the human accounts shown here may not reflect the true distribution of a random sample of Twitter accounts. These users have retweeted a politician's post and are more engaged on the platform. For a comparison of bots against a true random sample of any Twitter users, please look here.

F Bot Percentage binned by quantiles and follower numbers

This is the equivalent of Figure 6 shown in the paper where Y axis represents the number of normalised followers and the bot percentage is binned by quantiles.

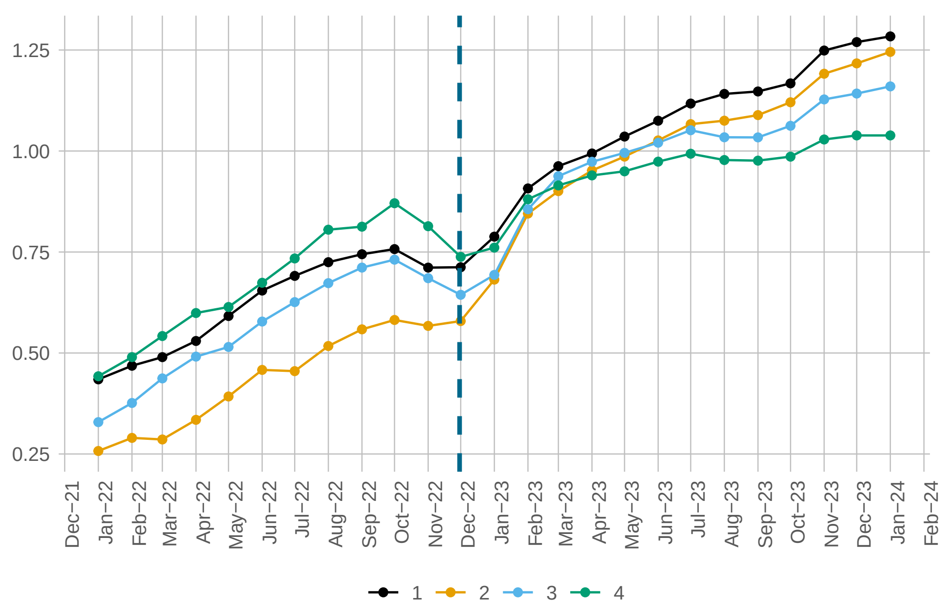
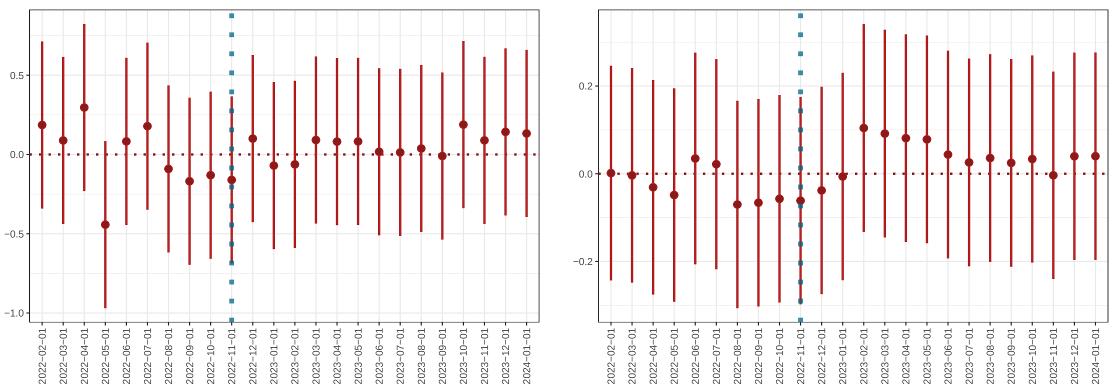


Figure 13: Caption

G Event Study Output plots for Facebook and Instagram

Figure 14: Parallel trends in the normalized follower counts before the intervention



Note: The graph shows the results of the event study analysis. The coefficients show no disruption on Facebook likes and Instagram follower counts post-shutdown for politicians with high bot usage. The interaction coefficients are shown above. The blue dotted line is the first reported day of the global outage. The policy was officially announced in Feb 2023.

H Summary statistics of online news articles

	N	Mean	SD	Median	Min	Max
Jan 2022	535	5.1	15.7	1.3	0	190
Feb 2022	535	4.4	9.7	1.4	0	97
Mar 2022	535	5.1	12.1	1.6	0	150
Apr 2022	535	4.5	10.4	1.6	0	105
May 2022	535	6.2	15.1	1.8	0	213
Jun 2022	535	6.8	16.0	2.1	0	169
Jul 2022	535	5.5	14.9	1.7	0	201
Aug 2022	535	6.7	31.1	1.5	0	646
Sep 2022	535	4.0	11.6	1.2	0	178
Oct 2022	535	4.6	14.2	1.2	0	240
Nov 2022	535	7.8	24.9	2.0	0	405
Dec 2022	535	5.7	16.3	1.5	0	213
Jan 2023	535	7.6	30.1	1.8	0	584
Feb 2023	535	5.3	16.3	1.6	0	267
Mar 2023	535	4.9	14.3	1.5	0	240
Apr 2023	535	4.6	15.5	1.2	0	255
May 2023	535	4.7	20.0	1.1	0	417
Jun 2023	535	3.9	11.6	1.1	0	197
Jul 2023	535	3.3	7.6	1.1	0	78
Aug 2023	535	2.6	7.2	0.8	0	90
Sep 2023	535	4.4	14.5	1.0	0	238
Oct 2023	535	7.6	30.0	1.4	0	488
Nov 2023	535	4.0	11.8	1.0	0	165
Dec 2023	535	3.6	8.9	0.9	0	117
Jan 2024	535	3.7	10.8	0.9	0	142
Feb 2024	535	6.8	23.4	1.3	0	381

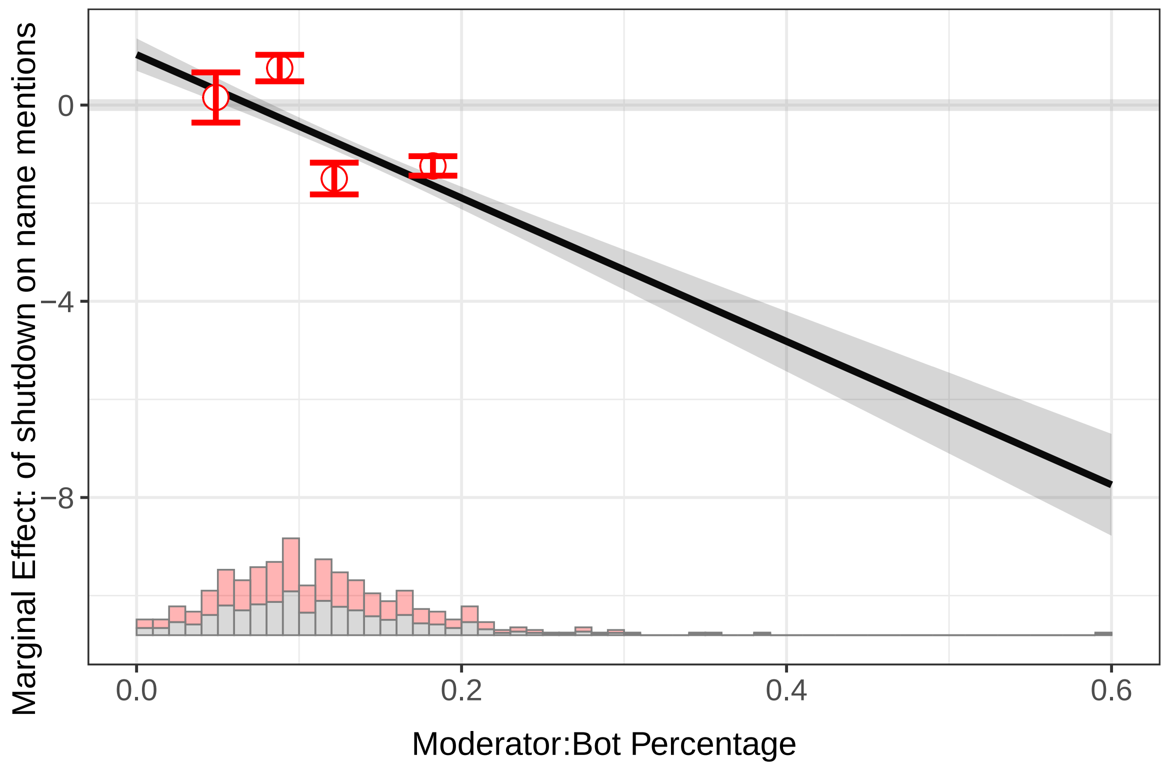
Table 7: Daily Online News articles that mention a politicians name

I Sample of Online News Coverage for Matt Gaetz

URL	Date	Title
https://www.politicususa.com/2023/11/21/even-florida-cant-stand-matt-gaetz.html	2023-11-21 16:15:00	Even Florida Cant Stand Matt Gaetz
https://radaronline.com/p/kevin-mccarthy-psychotic-matt-gaetz-bittersweet-house-exit/	2023-12-15 19:30:00	Kevin McCarthy Trashes Matt Gaetz Ahead of House Exit
https://radaronline.com/p/matt-gaetz-trolled-by-fake-award-trafficking-allegations/	2023-12-13 01:30:00	Rep . Matt Gaetz Falls For Troll Fake Award at Christmas Event
https://crooksandliars.com/2024/02/oops-ethics-committee-has-matt-gaetzs-sex	2024-02-16 15:30:00	Oops ! Ethics Committee Has Matt Gaetz Sex - Text Messages
https://hotair.com/headlines/2023/12/30/the-top-5-political-surprises-of-2023-n601865	2023-12-30 23:45:00	The Top 5 Political Surprises Of 2023
https://freerepublic.com/focus/f-news/4197999/posts	2023-11-19 23:15:00	McCarthy : Gaetz Could Have Same Problem as Santos when Ethics Complaint Comes Out
https://www.dailymail.co.uk/news/article-12775489/Matt-Gaetzs-approval-Florida-plummets-21-led-bid-oust-Kevin-McCarthy-Speaker-results-ethics-investigation-linger.html	2023-11-21 16:15:00	Matt Gaetz approval in Florida plummets to 21 % after he led the bid to oust Kevin McCarthy as Speaker and as the results of his ethics investigation linger
https://www.mediaite.com/news/living-off-your-daddys-money-republican-senator-roasts-matt-gaetz-in-vicious-spat-over-stock-trades/	2023-12-15 23:45:00	Matt Gaetz , Markwayne Mullin Spar Over Stock Trades On X
https://politicalwire.com/2024/02/14/house-panel-obtains-matt-gaetzs-texts/	2024-02-14 22:30:00	House Panel Obtains Matt Gaetz Texts
https://www.foxnews.com/politics/conservative-firebrand-praises-matt-rosendale-ahead-potential-senate-bid-shock-system	2024-01-25 23:45:00	Conservative firebrand praises Matt Rosendale ahead of potential Senate bid : Shock to the system
https://www.yahoo.com/entertainment/matt-gaetz-house-gop-token-183032877.html	2024-01-24 19:45:00	Matt Gaetz , the House GOP Token Troublemaker , Caught in Escalating Ethics Probe That Could Lead to Expulsion
https://www.thedailybeast.com/matt-gaetz-invites-julios-and-jamals-to-replace-maga-karens	2024-01-18 21:30:00	Matt Gaetz Invites Julio and Jamal to Replace MAGA Karens
https://news.yahoo.com/matt-gaetz-know-why-congressional-160044225.html	2023-12-08 16:45:00	Matt Gaetz and what we know about why his congressional colleagues seem to detest him

J Marginal Effects at different quantiles of bot percentage

This was calculated using Hainmueller et al. (2019) methodology and the binning estimator.



K Summary statistics of T.V name mentions

Channel	Mean	Median	SD	Min	Max
BLOOMBERG	0.01	0	0.64	0	245
FBC	0.09	0	0.88	0	128
CNBC	0.02	0	0.37	0	101
CSPAN	0.18	0	1.98	0	305
CSPAN2	0.07	0	0.83	0	132
CSPAN3	0.04	0	0.70	0	172
CNN	0.17	0	2.46	0	366
FOXNEWS	0.23	0	2.25	0	307
MSNBC	0.27	0	2.94	0	407

L Robustness of results at different cutoff values

L.1 Digital News

Table 8: Robustness of results at various cutoff values

	Dependent variable:Average Daily Articles							
	Cutoff:2.5		Cutoff:3.0		Cutoff:3.5		Cutoff:4.0	
	(1)	(2)		(2)	(1)	(2)	(1)	(2)
After	−0.412	−0.668***	0.193	−0.532***	0.732***	−0.335***	1.029***	−0.016
	(0.272)	(0.111)	(0.270)	(0.111)	(0.239)	(0.107)	(0.176)	(0.096)
Bot%×After	−0.641	0.049	−2.653***	−0.551	−5.919***	−1.915***	−14.616***	−7.765***
	(0.776)	(0.370)	(0.827)	(0.397)	(0.965)	(0.507)	(1.334)	(0.858)
Observations	312,756	428,532	312,756	428,532	311,952	428,532	310,344	428,532
Adjusted R^2	0.254	0.255	0.254	0.255	0.254	0.255	0.254	0.255

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 9: Robustness of results at different cutoff values (4.1-4.5)

	Dependent variable: Value									
	Cutoff:4.1		Cutoff:4.2		Cutoff:4.3		Cutoff:4.4		Cutoff:4.5	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
After	1.101***	0.075	1.044***	0.093	1.039***	0.138	1.131***	0.212**	1.028***	0.186**
	(0.164)	(0.093)	(0.156)	(0.091)	(0.148)	(0.089)	(0.143)	(0.087)	(0.139)	(0.086)
Bot%×After	−19.565***	−11.398***	−22.521***	−13.897***	−25.608***	−16.808***	−41.842***	−28.441***	−45.092***	−31.572***
	(1.552)	(1.038)	(1.720)	(1.182)	(1.805)	(1.277)	(2.658)	(1.912)	(2.899)	(2.115)
Observations	309,540	428,532	309,540	428,532	309,540	428,532	309,540	428,532	309,540	428,532
Adjusted R^2	0.254	0.255	0.254	0.255	0.254	0.255	0.255	0.255	0.255	0.255

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

L.2 TV News coverage

Table 10: Robustness to cutoff TV Results (2.5-4.0)

	CNBC	FBC	BLOOMBERG	CSPAN	CSPAN2	CSPAN3	FOXNEWS	MSNBC	CNN
Cutoff:2.5									
	Dependent variable:Counts of name mentions on TV channel								
After	0.004	0.010	0.004	0.078**	0.045***	−0.009	0.024	−0.037	0.041
	(0.006)	(0.013)	(0.010)	(0.030)	(0.013)	(0.011)	(0.033)	(0.045)	(0.038)
Bot%×Post	−0.004	−0.015	−0.005	−0.033	−0.027**	0.004	−0.084**	−0.046	−0.031
	(0.006)	(0.014)	(0.011)	(0.033)	(0.014)	(0.012)	(0.036)	(0.049)	(0.042)
Observations	421,083	421,083	421,083	421,083	421,083	421,083	421,083	421,083	421,083
Adjusted R^2	0.041	0.139	0.037	0.102	0.116	0.073	0.170	0.103	0.069
Cutoff:3.0									
	CNBC	FBC	BLOOMBERG	CSPAN	CSPAN2	CSPAN3	FOXNEWS	MSNBC	CNN
After	0.005	0.013	0.004	0.083***	0.046***	−0.009	0.033	−0.027	0.047
	(0.006)	(0.013)	(0.010)	(0.030)	(0.013)	(0.011)	(0.033)	(0.045)	(0.038)
Bot%×Post	−0.009	−0.030**	−0.007	−0.057	−0.033**	0.005	−0.131***	−0.094*	−0.064
	(0.007)	(0.015)	(0.012)	(0.035)	(0.015)	(0.013)	(0.039)	(0.052)	(0.045)
Observations	421,083	421,083	421,083	421,083	421,083	421,083	421,083	421,083	421,083
Adjusted R^2	0.041	0.139	0.037	0.102	0.116	0.073	0.170	0.103	0.069
Cutoff:3.5									
	CNBC	FBC	BLOOMBERG	CSPAN	CSPAN2	CSPAN3	FOXNEWS	MSNBC	CNN
After	0.006	0.018	0.004	0.089***	0.046***	−0.008	0.047	−0.013	0.060
	(0.006)	(0.013)	(0.010)	(0.030)	(0.013)	(0.011)	(0.033)	(0.045)	(0.038)
Bot%×Post	−0.018**	−0.066***	−0.011	−0.116**	−0.049***	0.003	−0.261***	−0.208***	−0.159***
	(0.009)	(0.020)	(0.015)	(0.045)	(0.019)	(0.016)	(0.049)	(0.067)	(0.057)
Observations	421,083	421,083	421,083	421,083	421,083	421,083	421,083	421,083	421,083
Adjusted R^2	0.041	0.139	0.037	0.102	0.116	0.073	0.170	0.103	0.069
Cutoff:4.0									
	CNBC	FBC	BLOOMBERG	CSPAN	CSPAN2	CSPAN3	FOXNEWS	MSNBC	CNN
After	0.007	0.022*	0.004	0.101***	0.045***	−0.008	0.072**	0.009	0.082**
	(0.006)	(0.013)	(0.010)	(0.030)	(0.013)	(0.011)	(0.033)	(0.044)	(0.038)
Bot%×Post	−0.053***	−0.192***	−0.016	−0.375***	−0.081**	−0.001	−0.829***	−0.688***	−0.599***
	(0.015)	(0.033)	(0.025)	(0.076)	(0.032)	(0.027)	(0.084)	(0.113)	(0.097)
Observations	421,083	421,083	421,083	421,083	421,083	421,083	421,083	421,083	421,083
Adjusted R^2	0.041	0.139	0.037	0.102	0.116	0.073	0.171	0.104	0.069

Note: *p<0.1; **p<0.05; ***p<0.01

Table 11: Robustness to cutoff TV Results (4.1-4.5)

	CNBC	FBC	BLOOMBERG	CSPAN	CSPAN2	CSPAN3	FOXNEWS	MSNBC	CNN
Cutoff:4.1									
	Dependent variable:Counts of name mentions on TV channel								
After	0.008	0.024^{*}	0.004	0.105^{***}	0.045^{***}	-0.007	0.075^{**}	0.017	0.088^{**}
	(0.006)	(0.013)	(0.010)	(0.030)	(0.012)	(0.011)	(0.033)	(0.044)	(0.038)
Bot%×Post	-0.078^{***}	-0.265^{***}	-0.024	-0.547^{***}	-0.111^{***}	-0.005	-1.110^{***}	-1.005^{***}	-0.851^{***}
	(0.018)	(0.040)	(0.031)	(0.093)	(0.039)	(0.033)	(0.101)	(0.137)	(0.117)
Observations	421,083	421,083	421,083	421,083	421,083	421,083	421,083	421,083	421,083
Adjusted R^2	0.041	0.139	0.037	0.102	0.116	0.073	0.171	0.104	0.069
Cutoff:4.2									
	CNBC	FBC	BLOOMBERG	CSPAN	CSPAN2	CSPAN3	FOXNEWS	MSNBC	CNN
After	0.008	0.025^{*}	0.004	0.106^{***}	0.045^{***}	-0.007	0.076^{**}	0.019	0.089^{**}
	(0.006)	(0.013)	(0.010)	(0.030)	(0.012)	(0.011)	(0.033)	(0.044)	(0.038)
Bot%×Post	-0.096^{***}	-0.340^{***}	-0.034	-0.664^{***}	-0.134^{***}	-0.006	-1.342^{***}	-1.242^{***}	-1.043^{***}
	(0.020)	(0.046)	(0.035)	(0.105)	(0.044)	(0.038)	(0.115)	(0.156)	(0.133)
Observations	421,083	421,083	421,083	421,083	421,083	421,083	421,083	421,083	421,083
Adjusted R^2	0.041	0.139	0.037	0.102	0.116	0.073	0.171	0.104	0.069
Cutoff:4.3									
	CNBC	FBC	BLOOMBERG	CSPAN	CSPAN2	CSPAN3	FOXNEWS	MSNBC	CNN
After	0.008	0.025^{*}	0.004	0.107^{***}	0.045^{***}	-0.007	0.076^{**}	0.020	0.089^{**}
	(0.006)	(0.013)	(0.010)	(0.030)	(0.012)	(0.011)	(0.033)	(0.044)	(0.038)
Bot%×Post	-0.113^{***}	-0.382^{***}	-0.038	-0.786^{***}	-0.146^{***}	-0.003	-1.531^{***}	-1.422^{***}	-1.187^{***}
	(0.022)	(0.049)	(0.038)	(0.114)	(0.048)	(0.041)	(0.124)	(0.169)	(0.144)
Observations	421,083	421,083	421,083	421,083	421,083	421,083	421,083	421,083	421,083
Adjusted R^2	0.041	0.139	0.037	0.102	0.116	0.073	0.171	0.104	0.069
Cutoff:4.4									
	CNBC	FBC	BLOOMBERG	CSPAN	CSPAN2	CSPAN3	FOXNEWS	MSNBC	CNN
After	0.009	0.025^{*}	0.004	0.111^{***}	0.043^{***}	-0.007	0.083^{**}	0.026	0.093^{**}
	(0.006)	(0.013)	(0.010)	(0.030)	(0.012)	(0.011)	(0.033)	(0.044)	(0.038)
Bot%×Post	-0.193^{***}	-0.601^{***}	-0.051	-1.353^{***}	-0.157^{**}	-0.011	-2.578^{***}	-2.398^{***}	-1.950^{***}
	(0.033)	(0.074)	(0.057)	(0.171)	(0.071)	(0.061)	(0.187)	(0.253)	(0.216)
Observations	421,083	421,083	421,083	421,083	421,083	421,083	421,083	421,083	421,083
Adjusted R^2	0.041	0.139	0.037	0.102	0.116	0.073	0.171	0.104	0.069
Cutoff:4.5									
	CNBC	FBC	BLOOMBERG	CSPAN	CSPAN2	CSPAN3	FOXNEWS	MSNBC	CNN
After	0.008	0.024^{*}	0.004	0.109^{***}	0.043^{***}	-0.007	0.079^{**}	0.021	0.089^{**}
	(0.006)	(0.013)	(0.010)	(0.030)	(0.012)	(0.011)	(0.032)	(0.044)	(0.038)
Bot%×Post	-0.202^{***}	-0.654^{***}	-0.045	-1.456^{***}	-0.185^{**}	-0.008	-2.803^{***}	-2.588^{***}	-2.111^{***}
	(0.037)	(0.082)	(0.063)	(0.189)	(0.079)	(0.067)	(0.206)	(0.279)	(0.239)
Observations	421,083	421,083	421,083	421,083	421,083	421,083	421,083	421,083	421,083
Adjusted R^2	0.041	0.139	0.037	0.102	0.116	0.073	0.171	0.104	0.069

Note: *p<0.1; **p<0.05; ***p<0.01

M Online Supplemental Material

Additional Material will be posted here.