Proceedings of the Nineteenth International AAAI Conference on Web and Social Media (ICWSM 2025)

Election Polls on Social Media: Prevalence, Biases, and Voter Fraud Beliefs

Stephen Scarano[1], Vijayalakshmi Vasudevan[1], Mattia Samory[2], Kai-Cheng Yang[3], JungHwan Yang[4], Przemyslaw A. Grabowicz[1,5]

[1] University of Massachusetts Amherst, USA
[2] Sapienza University of Rome, IT
[3] Northeastern University, USA
[4] University of Illinois Urbana-Champaign, USA
[5] University College Dublin, IE
{sscarano, vijayalakshm}@umass.edu, mattia.samory@uniroma1.it, ka.yang@northeastern.edu, junghwan@illinois.edu, grabowicz@cs.umass.edu

Abstract

Social media platforms allow users to create polls to gather public opinion on diverse topics. However, we know little about what such polls are used for and how reliable they are, especially in significant contexts like elections. Focusing on the 2020 presidential elections in the U.S., this study shows that outcomes of election polls on Twitter deviate from election results despite their prevalence. Leveraging demographic inference and statistical analysis, we find that Twitter polls are disproportionately authored by male Republicans and exhibit a large bias towards candidate Donald Trump in comparison to mainstream polls. We investigate potential sources of biased outcomes from the point of view of inauthentic, automated, and counter-normative behavior. Using social media experiments and interviews with poll authors, we identify inconsistencies between public vote counts and those privately visible to poll authors, with the gap potentially attributable to purchased votes. We find that election polls tend to be more biased, contain more questionable votes, and attract more bots *before* the election day than *after*. We highlight and compare key factors contributing to biased poll outcomes. Finally, we identify instances of polls spreading voter fraud conspiracy theories and estimate that a couple of thousand such polls were posted in 2020. The study discusses the implications of biased election polls in the context of transparency and accountability of social media platforms.

Introduction

The advent of social media and the Internet has revolutionized the landscape of political discourse. The unique technological affordances provided by popular social media platforms, such as *Twitter*[1] (now known as *X*) and *Facebook*, enable people to become commentators and active distributors of information rather than mere consumers of it. Consequently, social media have garnered significant attention from the public, media, and political elites, who now use these online discussions to gauge the public's views on critical issues and utilize those insights for political campaigning (Harfoush 2009; Hughes et al. 2010).

However, social media may not accurately represent public opinion. Instead, it disproportionately reflects the views

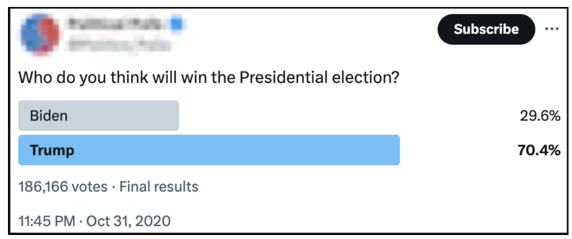

Figure 1: An example of a social poll from our dataset.

expressed by a reactive, polarized audience (Zhang, Chen, and Rohe 2022). Additionally, the demographic profile of Twitter users is non-representative (Wojcik and Hughes 2019), and the presence of bots, astroturfing accounts (Keller et al. 2020; Ferrara et al. 2016), and foreign influence (Schoch et al. 2022) further complicates the situation. The overrepresentation of certain perspectives, driven by a minority of users, can skew social media users' perceptions of reality and bias political processes.

Public opinion research has shown that opinion polls can influence how people perceive public opinion and shape how they form their own opinions (Lang and Lang 1984; Marsh 1985; Morwitz and Pluzinski 1996). Politicians and news outlets may also place greater emphasis on the issues that the public cares more about, further shaping policy decisions (Arnesen et al. 2017). Given the significance of polling in policy-making, it is important to note that Twitter, a leading social media platform, introduced polls in 2015,[2] a year before the 2016 U.S. presidential elections.

Social polls are straw polls that any social media user can create and share with others (see example in Figure 1). As a social media feature, they have become remarkably successful. In 2022, Elon Musk notably used this feature to make key business decisions, such as changing Twitter's CEO (Mehta 2022). Twitter polls have gained popularity due to their wide reach, quick turnaround time, ease of use, and low cost. It is common to see popular Twitter polls that amass millions of votes. However, no research has yet examined the extent to which social polls are used in political cam-

[1]Throughout the paper, we will use *Twitter* to refer to *X*.

[2]https://blog.twitter.com/official/en_us/a/2015/introducing-twitter-polls.html

paigns and the role they play. To bridge this research gap, our first goal is to address the following research question.

RQ1: *How many Twitter polls related to the U.S. presidential election were published in 2020? Did their number grow between the two election cycles, from 2016 to 2020?*

Twitter polls inherently lack scientific rigor due to the absence of systematic sampling and demographic information about the respondents, resulting in potential bias in poll outcomes. Despite this, little attention has been paid to understanding the biases in social polls and their broader impact. Therefore, we study the biases in social polls, focusing on the 2020 U.S. presidential election, to address the following set of related research questions.

RQ2.1: *How do Twitter election poll outcomes deviate from election results and traditional poll outcomes?*

After establishing that Twitter poll outcomes are biased, we attempt to reason about the potential sources of these biases. First, we show that Twitter poll votes can be purchased from online vendors and, then, we address the following research question.

RQ2.2: *How does Twitter account for purchased votes? Did such questionable votes appear in polls related to the 2020 U.S. presidential election?*

Second, election polls may have suspicious participants, such as bots, foreign accounts, or hyperactive users. We compare the percentage of Twitter users that are identified as automated accounts with a reference set, before and after elections, to address the following research question.

RQ2.3: *Was the activity of bots, foreign accounts, and hyperactive users larger – particularly before the election – for election polls than for comparable other polls?*

Third, poll outcomes may also be influenced by legitimate demographic and political differences among users who participate in the polls, which leads to the following question.

RQ2.4: *What are the characteristics of users engaging with Twitter election polls in comparison to the 2020 U.S. presidential election voters?*

Finally, we compare these three factors and study their relationship with biases in poll outcomes to address the following question.

RQ2.5: *Which attributes of users significantly correlate with biases in Twitter election poll outcomes?*

Individuals wary of institutions may perceive social polls as more legitimate and correct measures of public opinion than mainstream polls, in that the former are borne by the initiative of any social media user. This perception is especially problematic, considering that one of the major issues during the 2020 U.S. presidential elections was the allegation of voter fraud by presidential candidate Donald Trump and his supporters (Pennycook and Rand 2021). The issue has remained so prevalent that, as of the writing of this manuscript, it is still influencing the discourse surrounding the 2024 U.S. presidential election. At the core of the voter fraud conspiracy theory is the belief that the official election results do not represent public opinion. We focus on Twitter polls as a potential source of such beliefs.

In particular, while biases in the outcomes of Twitter election polls may misrepresent the popular support for presidential candidates, it is not clear whether such biased polls are indeed *explicitly* and *intentionally* used to support voter fraud beliefs surrounding the 2020 U.S. presidential election. Here, we address the following research questions.

RQ3: *How many Twitter election polls explicitly question election integrity and the accuracy of mainstream polls? How many users interact with such conspiratorial polls?*

We conclude the study by discussing the broader implications of biased election polls, including transparency and accountability of social media platforms such as Twitter.

Related Work

We analyze thousands of Twitter polls related to U.S. presidential elections, focusing on astroturfing and voter fraud conspiracy. To our knowledge, such polls have not been studied, except for our parallel work with a different focus on general statistical description (Scarano et al. 2024).

Potential Biases in Social Polls Public opinion, defined as an aggregate of individual opinions (Price 1992), is essential for democracy. During the election, understanding public opinion becomes particularly important as it offers insight into the voters' support towards political candidates, which can inform campaign strategies. One popular way to assess public opinion is through survey-based opinion polls (Price 1992). To accurately measure public opinion using polls, it is crucial to draw an unbiased and representative sample of the target population (i.e., electorates) (Squire 1988).

Over the past decade, social media have emerged as significant platforms for individuals to express their opinions on various social issues. However, opinions on social media may not accurately represent public opinion despite their sheer volume. This discrepancy arises due to several factors. First, the opinions of specific demographics can be over-represented because (i) Twitter users are more likely to be male and young (Mislove et al. 2011; Wojcik and Hughes 2019), (ii) nearly 80% of tweets are published by the 10% most active users, and (iii) politically-interested users are non-representative of all users (Hughes and Asheer 2019; Hughes 2021). Additionally, the prevalence of bot accounts, astroturfing campaigns, and artificial likes and comments can distort various metrics from Twitter, including social polls (Keller et al. 2020; Ferrara et al. 2016).

In the following section, we discuss related work on two potential sources of bias: inauthentic behavior, which may make active attempts to manipulate the visibility of certain views, and the lack of representativeness of the users who engage in social polling compared to the actual population of voters, which may over-emphasize specific political views.

Digital Manipulation Inauthentic social media entities, such as bots and astroturfing accounts, can contribute to social media biases. This phenomenon was famously manifested in Russia's *Internet Research Agency* (IRA) intervention in the 2016 U.S. Presidential election, where many organized or sponsored inauthentic social media operations were uncovered and documented (Keller et al. 2020; Schoch et al. 2022). Similarly, research has found that numerous bot accounts generate a significant amount of social media posts (Ferrara et al. 2016). Furthermore, the prevalence of bot accounts on Twitter has increased since Elon Musk's acquisition of the company in 2022 (Hickey et al. 2023).

Tools for manipulating social media discourse, now broadly accessible via online services, also have expanded their reach beyond governmental institutions to everyday consumers (Al-Rawi and Rahman 2020). These external influences have the potential to distort the representation of public opinion, particularly as politically motivated campaigns deliberately exploit them to sway the course of social media dialogues (Schoch et al. 2022).

Inauthentic accounts can influence Twitter discourse beyond merely generating messages or retweeting. For instance, in the 2012 South Korean election, national intelligence agents conducted an extensive disinformation campaign supporting one of the presidential candidates by manually operating multiple Twitter accounts. The agents tried to make these accounts appear genuine while amplifying messages through retweets, likes, and cross-posting (Keller et al. 2020). Similarly, the `#YaMeCanse` online protest movement in Mexico demonstrated coordinated influence, where astroturfing accounts manipulated the hashtag to stifle discourse (Suárez-Serrato et al. 2016).

While prior research has primarily focused on analyzing social media messages and sharing patterns (Keller et al. 2020; Schoch et al. 2022; Suárez-Serrato et al. 2016; Zhang, Chen, and Rohe 2022; Ferrara et al. 2016), it has largely overlooked potential social polls and their potential manipulation. This study uniquely contributes to the field by examining inauthentic activities and biases in social polls.

Demographics of Social Media Users One of the main sources of biases in social polls is the unrepresentative demographic traits of social media users. In particular, social media users do not accurately reflect the demographic makeup and location distribution of the general population. For instance, audiences on Twitter and Facebook are younger and lean more toward the political left (Mellon and Prosser 2017; Wojcik and Hughes 2019). Facebook users skew towards a female demographic, while Twitter users towards male (Mislove et al. 2011; Mellon and Prosser 2017; Wojcik and Hughes 2019). In addition, social media users generally tend to over-represent the wealthier and more educated population (Wojcik and Hughes 2019).

Another important factor contributing to potential biases in social polls is the political interest and ideology of social media users. According to data from the Pew Research Center, social media users in the United States tend to skew more liberal in their political affiliations compared to the general population. For instance, 36% identify as Democrats among social media users, whereas the corresponding figure in the U.S. general population is 30% (Wojcik and Hughes 2019).

Conspiracy Beliefs via Polls Opinion polls on social media are often driven by the prevalence of conspiratorial beliefs, such as voter fraud beliefs, in contemporary politics. Research has found that beliefs regarding both mail-in voter fraud and the manipulation of social media were popular during U.S. elections (Benkler et al. 2020; Ferrara et al. 2020). These studies highlight how such misinformation has the potential to significantly undermine the integrity and proper functioning of democratic systems.

Year	**2016**	**2020**	
Dataset source	**Query**	**Query**	**Decahose**
Utilized to address	**RQ1**	**RQ1, 2**	**RQ1, 2.3, 3**
Polls related to elections	1,759	4,900	12,990
Polls gauging support	510	1,440	
Engagement counts	+	+	+
User demographics	Not shown	+	+
User political affiliation	Not shown	+	
Poll author survey		+	

Table 1: Overview of our three Twitter poll datasets and their use in addressing our Research Questions (RQs). The table highlights data gaps in the 2020 Decahose and 2016 datasets. These limitations stem from Twitter's discontinuation of academic API access.

Survey questions that incorporate conspiracy theories can inadvertently contribute to mistrust in mainstream information (Clifford and Sullivan 2023). This effect is largely attributed to the "question wording" and "panel conditioning," a phenomenon where exposure to specific ideas within surveys shapes respondents' beliefs (Kalton and Schuman 1982; Das, Toepoel, and van Soest 2011). This is particularly evident when similar questions are repeated in multiple surveys, where respondents tend to learn from the survey questions they have previously answered.

Polls that express conspiracy beliefs might be particularly influential because they often provide simplified, alternative explanations for complex events. Their impact is more pronounced when respondents have limited information (Marchlewska, Cichocka, and Kossowska 2018), hold strong pre-existing political beliefs (Pennycook, Cannon, and Rand 2018), or have a predisposition towards conspiratorial thinking (Uscinski, Klofstad, and Atkinson 2016). Thus, given the prominence of conspiratorial narratives around election integrity, the design and content of survey questions, especially those involving conspiracy theories about the electoral process or legitimacy of information, can promote mistrust in the government and the overall democratic process (Pennycook and Rand 2021). However, the extent to which such polls exist on Twitter is largely unknown.

Datasets and Twitter Restrictions

We collected three large datasets of Twitter polls related to the 2020 and 2016 U.S. presidential elections (see Table 1). All datasets contain polls that explicitly include both major presidential candidates as poll options. To estimate the growth and overall prevalence of election polls (RQ1), we use the dataset for 2016 and the two datasets for 2020. For other research questions, we use only the 2020 datasets, since the restrictions of the Twitter API prevented us from collecting a more complete 2016 dataset. Below, we describe the datasets and the implications of Twitter API restrictions.

To facilitate the reproduction of the results of this study, we share the tweet IDs of all collected polls on GitHub.[3]

[3]https://github.com/social-info-lab/twitter_us-election-

Datasets

Election Query Polls Provided access to the Twitter API v2, we collected election polls from the periods of 1/1/2020–11/30/2020 and 1/1/2016–11/30/2016 by making full-archive searches for the queries "`vote AND (trump OR biden)`" and "`vote AND (trump OR clinton)`" respectively. To identify polls related to presidential elections, we focused on the polls that explicitly mention respective presidential candidates among poll options ("`trump OR donald`" and "`joe OR biden`" or "`hillary OR clinton`"). This resulted in 4,900 polls. We refer to these polls as query polls to distinguish them from random polls described next.

We analyze the 2016 and 2020 query datasets in more detail in our parallel work (Scarano et al. 2024), whereas here we study a larger dataset of polls, described next, focusing on their potential manipulation.

Random Election Polls We leveraged the Decahose stream, a 10% random sample of all tweets produced on Twitter, to identify random election polls. We excluded from this dataset all retweets. Without this filtering, our sample would be biased towards popular polls and would not be representative of all polls, leading to incorrect estimates. Then, to identify polls related to elections, we applied the same filtering step as for the query polls, i.e., we select polls that mention "`trump`" and "`biden`" among poll options.

The main advantage of this dataset is that it is a random 10% sample (without retweets) of all polls, whereas the query dataset includes only the polls that mention the word "vote" and either "trump" or "biden" in its main text (not poll options), and we do not know what fraction of all polls they represent. However, the Decahose sample of polls has a few crucial limitations. First, the Decahose stream does not contain any polls that were published in 2016, likely because the polling feature was introduced to Twitter just one year earlier. Second, due to Twitter API restrictions, we are unable to complete it, as we explain in the next subsection.

Mainstream and Exit Polls To make comparisons with the Twitter polls, we use 192 mainstream polls from 2020 aggregated by *FiveThirtyEight*,[4] and national exit poll data distributed by the *Roper Center* (NBC) 2020).

Polls Gauging Support for Presidential Candidates To compare the results of Twitter election polls with election results and mainstream polls (RQ2.1) and to explain biases in Twitter poll outcomes (RQ2.5), we must identify Twitter polls that gauge popular support for the U.S. presidential candidates. For instance, polls that ask "Who won the last debate?" are related to the elections and appear in our datasets but do not gauge the overall (rather than momentary) support for the presidential candidates. To identify relevant polls, we manually inspected all of the query polls from 2016 and 2020. To this end, we developed a labeling guideline that defines polls gauging support for presidential candidates as either (i) directly asking for voting preferences (e.g., "Who has your vote?"), or (ii) asking for election predictions (e.g., "Who do you think will win the presidential election?"). This approach identified a total of 1,440 Twitter polls gauging support for the 2020 presidential candidates. To estimate the inter-rated agreement, a subset of 194 polls was labeled by two trained coders. They have achieved an almost perfect inter-rater agreement on this set of items, measured with Cohen's kappa, of 0.914 ($p < 0.001$).

Classification and Validation of User Traits For each poll from our three datasets, we collect all retweeters and favoriters. For the polls from the query-based datasets, we collect followees of all poll authors, necessary for political affiliation identification. And when possible, we use state-of-the-art classifiers to generate the bot score, age, gender, organization status, and political affiliation for each poll author, retweeter, and favoriter. We describe the classifiers in the respective method sections focused on these user traits.

To validate the classifier labels, we asked one human coder to review the estimated outcome of those four attributes from a random set of 239 Twitter accounts participating in the vote query polls. To simplify the process, we showed the machine-determined attributes to the coder and asked them to determine whether they agreed with those classifications or not. Based on the coder's validation, our methods achieved approximately 93% accuracy in distinguishing between organizational and personal accounts, 91% accuracy in assessing bot-likeness, 93% accuracy in estimating political ideology, 91% accuracy in estimating the age of the account holders, and 88% accuracy in classifying the gender of the account holders.

Use and Limitations of Datasets

Impact of Twitter API Restrictions Due to the closure of Academic API by Twitter in 2023 and changes to the Twitter user interface at the end of 2022, we were not able to: (a) conduct a survey of poll authors neither for the 2016 query polls nor for the Decahose polls, and (b) collect the information about user followees necessary to infer political affiliations of users interacting with the Decahose polls (see Table 1 for summary). We further discuss these two issues in the sections addressing RQ2.2 and RQ2.4, respectively.

Datasets vs. Research Questions Given the varying strengths and limitations of different datasets, we use them for different research questions. The 2020 query dataset is our most complete dataset, and we use it as our main dataset. The Decahose dataset provides a random 10% sample of polls, which allows for estimating the total number of polls published on Twitter, so we use it to address RQ1 and RQ3. The Decahose dataset also offers many more samples than our query datasets, which is important when making precise temporal comparisons, so we use it together with the 2020 query dataset to address RQ2.3. Last but not least, the 2016 query dataset allows us to study the growth of social media polls (RQ1) by comparing it with the 2020 query dataset.

Rise of Election Polls on Twitter (RQ1)

We plot the number of polls, votes, poll retweets, poll favoriters, and the number of followers of the poll authors

polls_2016-2020

[4]https://projects.fivethirtyeight.com/polls

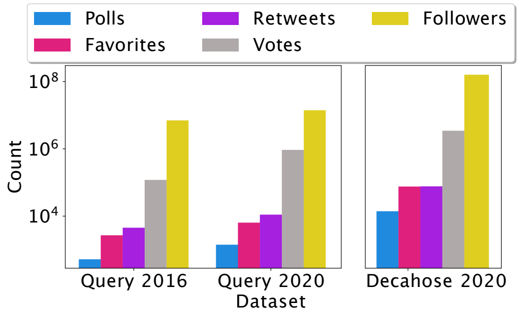

Figure 2: Growth of the number of polls, votes, poll retweets, poll favoriters, and followers of the poll authors. The difference between the numbers for the 2016 and 2020 query datasets (left and center) illustrates poll growth. The Decahose dataset (right) includes only about 10% of all polls, so the respective numbers represent *only* about one-tenth of all election polls posted on Twitter. The numbers suggest that election polls have grown in popularity on Twitter.

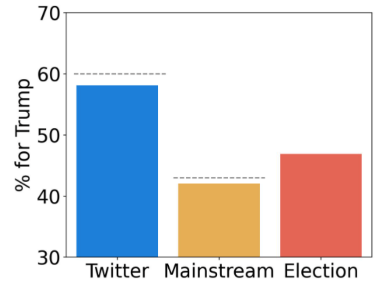

Figure 3: The average outcome of Twitter polls, the average outcome of mainstream polls, and the official election outcome. Dashed lines signify the median.

for each dataset in Figure 2. We note that all metrics have considerably grown between 2016 and 2020 (compare left and center of Figure 2), suggesting that election polls have grown in popularity on Twitter. If the growth trend between 2016 and 2020 extends into 2024, we expect the number of polls and votes to be even larger in 2024.

We note that the numbers from the Decahose dataset likely underestimate the true numbers of polls and users engaging with them on Twitter by a factor of 10. As mentioned earlier, Twitter Decahose provides a 10% sample of all tweets. To further confirm the percentage of polls included in our Decahose dataset, we leverage the dataset generated by Pfeffer et al. (2023), which contains all tweets from a single day. We use the same method to extract polls from this dataset, and confirm that our Decahose dataset contains about 10% of polls published that day. Since there were nearly 13,000 polls in Decahose in 2020, we estimate that about 130,000 polls related to the U.S. presidential candidates were posted on Twitter in 2020, and they attracted over 20 million votes. For reference, there were about 168.42 million registered voters in the 2020 U.S. presidential elections.

A caveat is that since the Decahose stream only includes the polls themselves, we could not determine the vote counts and outcomes of polls that were deleted or suspended.

Biases in Results of Election Polls (RQ2.1)

We compare the results of Twitter and mainstream polls to gauge support for the 2020 U.S. presidential candidates. Overall, the results of Twitter polls demonstrate that there is a substantial partisan slant, with a consistent leaning toward Trump (Figure 3). Specifically, the median support for Trump in social polls is 60% (average 58%), while in mainstream polls it is 43% (average 42%). While social media polls overestimate support for Trump compared to actual vote shares on election day, mainstream polls tend to underestimate Trump's support. We observe similar gaps in the 2016 election (Scarano et al. 2024). It is also worth mentioning that the variance of Twitter polls is substantially larger than that of mainstream surveying (750 vs. 10), suggesting that social poll outcomes are affected by a number of factors.

Focusing on the temporal trend shown in Figure 4, we observe a significant drop in support for Trump immediately after the election day. Several potential factors might contribute to this decline, such as digital astroturfing, differences in online behavior patterns between Trump and Biden supporters that wane post-election, and the influence of live election reporting. These findings shed light on the complex dynamics shaping social poll results and highlight the need for nuanced interpretation.

Partisan behavior may indeed differ between authors of Trump and Biden polls; in fact, there is reason to believe that Trump's coalition feels stronger about their support than that of Biden. The 2020 election polling measures a 20% gap in *strong* support for Trump's favor and, as a result, his voters may be more likely to post tweets, media, and polls related to his candidacy (Doherty et al. 2020). Political analysis on this front is out of scope for this paper, and instead, we seek to determine to what extent other factors might feasibly explain the gap. We turn to these questions next.

Questionable Votes in Social Polls (RQ2.2)

On June 19th, 2020, Polish state media *TVP INFO* ran an article detailing an abnormality in a poll it published on its Twitter account asking who won a Polish presidential debate (Zajaczkowski and Pereira 2020). *TVP INFO* claimed that 44.5% of the votes had been fraudulent, based on the difference between the private vote count reported in the tweet's analytics and the vote count visible to the public.

In this section, we corroborate this indicator of inorganic user behavior in Twitter polls: the discrepancy between the count of votes shown in public and privately to the poll author. We first verify that this discrepancy is ascribable to votes purchased on markets for inorganic user behavior. Then, we survey 984 authors of 1,440 polls from the 2020 query dataset for their private vote counts and estimate the fraction of questionable votes in the 2020 election polls as the discrepancy between public and private vote counts.

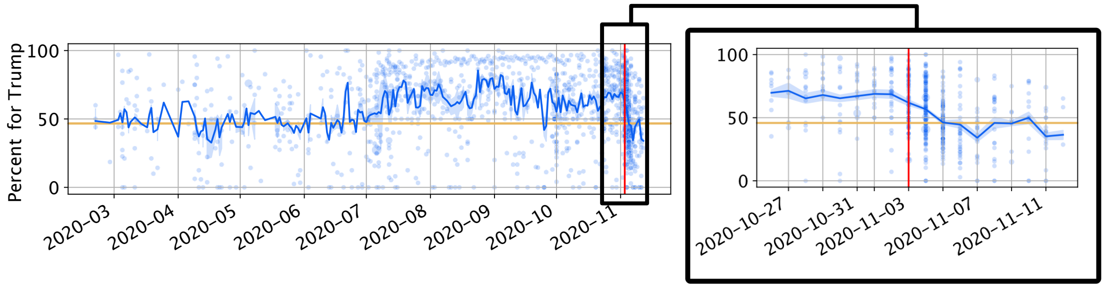

Figure 4: (Left) Moving average of the percent of votes for Donald Trump in Twitter polls published over the course of 2020. (Right) Zooming in on the period of the three weeks around the election day suggests that the bias towards Trump diminished right after the election day, marked with the vertical red bars. The election outcome (orange line) is shown for comparison.

Vote Count Discrepancy and Inauthentic Votes

Methods To emulate a potential agent seeking to distort an online poll, we searched on *Google* for the query "*buy Twitter poll votes*." Next, we randomly picked five of the top 10 online vendors (see full list in Appendix B).
Then, for each of the five vendors, we posted two identical polls and purchased 100 fake votes from the vendor per poll. Using a separate account, we submitted one organic vote for the candidate choice that was opposite to the inauthentic votes. We ran the polls for 24 hours, then tabulated and compared the publicly-listed and privately-listed vote counts found in the user analytics (Appendix C).

		True vote counts		Twitter counts	
Trial	Vendor of votes	Organic	Bought	Public	Priv.
1	viplikes.net	1	100	138	1
2	viplikes.net	1	100	120	1
3	socialboss.com	1	100	134	1
4	socialboss.org	1	100	125	1
5	socialwick.com	1	100	135	1
6	socialwick.com	1	100	126	1
7	gettwitterretweet.com	1	100	124	1
8	gettwitterretweet.com	1	100	120	1
9	instafollowers.com	1	100	120	1

Table 2: Discrepancy between public and private vote counts. Fake votes purchased in the experiment are included in the public vote count, but not in the private vote count.

Results Table 2 shows that—regardless of cause—Twitter's poll system records public votes that its private analytics does not acknowledge. Although only 100 votes were purchased in every case, vendors consistently provided more votes than that (between 120 and 140 votes), which likely corresponds to an attempt to avoid automatic detection. The private count of votes shows the correct number of organic votes provided on the polls, that is, one organic vote. In other words, none of the purchased votes counted towards the private count. We list potential explanations in the Discussion and Conclusions section.

Vote Count Discrepancy in Election Polls

Methods To estimate the number of questionable votes in our tweet poll dataset, we contacted 984 authors from the 2020 query dataset and requested the private vote count for their corresponding poll(s). It is no longer feasible to check private vote counts for older posts due to the changes in the Twitter user interface at the end of 2022. Hence, we could not conduct the same survey for the 2016 polls. Our survey and analysis of responses received IRB approval.

We sent each author a direct message or a tweet introducing the researchers (see Appendix A for the full message) before soliciting the private vote count of their respective poll(s). The message contained a link to the poll in question and instructions on retrieving the private vote count. It also clarified that the researchers would maintain the author's anonymity and only publish aggregate results.

To better interpret these results, we performed a placebo experiment. We posted a Twitter poll and solicited responses from our colleagues working at the college of the last author through an internal mailing list without explaining the purpose of this experiment. To the best of our knowledge, this poll contained human responses only, without the presence of inauthentic votes. We leverage the public/private vote count discrepancy measured in this placebo experiment to contextualize those of the surveyed poll authors.

Results We made multiple attempts to contact the poll authors, but most did not respond, probably because of the sensitive nature of the request (e.g., some of the users might have participated in political campaigns). In the end, we received responses from 22 poll authors. Although the number is not high, they offer critical insight into political polling on Twitter, consistent with the other parts of this study.

Of the 22 poll authors, 21 users reported discordance between public and private vote records, with a median 35% increase from private to public. We refer to the excessive part as *questionable votes*. Across all 22 polls, the total count of public votes is 3,605, while the private vote count is 1,555. Thus, 57% of all votes appear questionable. The fraction of questionable votes among different polls varies drastically, from 0% to 68%. The outcomes of these polls are mixed, with 60% supporting Trump on average, matching the overall average of 58% for our 2020 dataset.

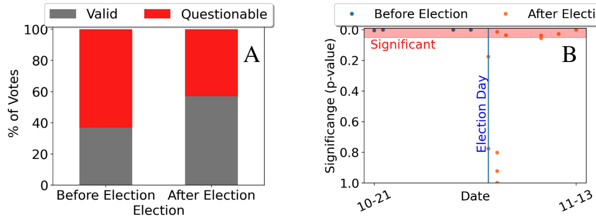

Figure 5: (A) The fraction of questionable votes drops after the election day. (B) The fraction of questionable votes is larger than expected for all evaluated polls before the election day and for a subset of polls after the election.

In our placebo experiment, where we believed all votes were organic, 33 out of 83 (40%) votes appeared questionable. Thus, some discrepancies between the public and private vote counts may be common and expected. For instance, it is possible that Twitter's classification of inauthentic votes is inaccurate and results in a fraction of "questionable"votes. However, even if we assume that 40% of public votes (as determined in the placebo experiment) are not counted as private votes due to legitimate reasons, we still observe potential evidence of manipulation.

First, the fraction of questionable votes is significantly higher ($p = 0.05$, Mann-Whitney U test) before than after the election day (Figure 5A). Second, for each poll, we test whether the fraction of questionable votes is significantly higher than the 40% of votes that may originate from legitimate sources. We conduct a binomial test comparing the observed and expected number of private votes, given that $p = 0.4$ and n is the public vote count. We find that 13 of the 21 polls exhibit a larger discrepancy between the public and private vote count than expected ($p = 0.05$, see Figure 5B). Furthermore, all polls posted *before* the election have significantly more questionable votes than we would expect under the assumption that 40% of public votes are not counted as private votes due to legitimate reasons.

Suspicious User Activity (RQ2.3)

In addition to scrutinizing polls, we analyze poll authors and participants for signals of inorganic activity. Here we focus on favoriters and retweeters of the polls since the identity of voters is not disclosed by Twitter. We determine whether users are bot-like, foreign, or hyperactive accounts. While bot activity has an intuitive relationship to manipulation, signals of foreign origin (non-U.S) and hyperactivity may be markers of coordinated astroturfing campaigns, foreign or otherwise. Here, we analyze the query and Decahose sets of 2020 election polls jointly.

Reference Set of Polls Matched to Election Polls To identify suspicious user activity, we compare bot activity, locations, and activity levels of users participating in election polls with those participating in random polls. We collect a set of random polls from 2020 (the vast majority non-political). For each election poll from our two 2020 datasets, we match a poll to one with a similar vote count from the reference set. And for each matched poll, we collect all its retweeters and favoriters for comparison. We refer to the matched poll set as *reference* polls.

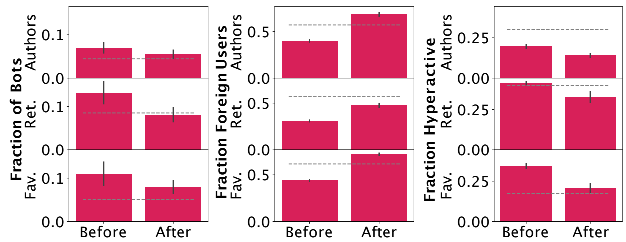

Figure 6: The fraction of (from left to right) bots, foreign accounts, and hyperactive users among (from top to bottom) poll authors, retweeters, favoriters before and after the election day (left and right bars within each panel). The dashed line corresponds to the respective fraction for random reference polls matched to the election polls based on the number of votes and publication year.

Bots

Methods We subjected authors, retweeters, and favoriters of 2020 election polls through the *Botometer-V4* classifier to estimate the distribution of bot accounts. *Botometer* is a random forest classifier that evaluates network, user, friend, temporal, content, and sentiment features to label a profile as authentic or artificial (Sayyadiharikandeh et al. 2020). To contextualize these results, we compared the outputs to a random sample of Twitter polls posted over the same period and of similar vote distribution. We then split user groups temporally–two weeks before and after the election–to identify elevated bot presence.

Results Figure 6 visualizes the percentage of bots among poll authors before and after the election in comparison to a random sample. The distributions for retweeters and favoriters follow the same qualitative patterns. Overall, we find that users participating in election polls have higher bot scores than users participating in the reference polls. Bot score distributions of poll authors differ significantly from our random sample ($p < 0.001$, MWU test), and significance also holds for retweeters and favoriters (see Appendix D for all significance values relevant to RQ2.3). We also find that the fraction of bots among retweeters and favoriters of political polls posted two weeks prior to election day are significantly higher than those two weeks afterward ($p < 0.001$ and $p < 0.001$, respectively; MWU tests).

Foreign Accounts

Methods On Twitter, users can optionally disclose their location in plain text using the `location` field of their profile. We resolve such entries to geolocations using Photon, an open-source geocoder.[5] For users whose locations could not be geocoded via the previous method, we combine the `location` and `description` plain-text fields

[5]https://photon.komoot.io

of the user profiles and extracted emojis corresponding to national flags, excluding the cases of users displaying flags of multiple countries. This allows us to distinguish between users belonging to foreign countries and the U.S.

Results Similar to bot accounts, the distributions of foreign accounts for authors, retweeters, and favoriters differ significantly from the respective random samples. However, occurrences of foreign accounts are *less* likely in political polls. Additionally, the prevalence of foreign authors is significantly higher ($p < 0.001$, MWU test) post-election in comparison to pre-election.

Hyperactive Users

Methods We refer to users who post, on average, more than 20 tweets per day as hyperactive. For all authors, retweeters, and favoriters, we estimate user activity as the average number of statuses (including retweets) since the creation date of the account.

Results Activity distributions for authors, retweeters, and favoriters differ significantly from their respective reference ($p < 0.001$ for all groups, MWU tests). Moreover, across the three groups, activity before the election is higher than post-election ($p < 0.001$ for all comparisons according MWU tests after adjusting for multiple comparisons). Figure 6 visualizes the percentage of hyperactive users before and after the election in comparison to a random sample.

Traits of Users Engaging with Polls (RQ2.4)

The biased outcomes of polls may be explained by factors besides suspicious activity in election polls, such as a skewed demographic and political composition of users involved in polling on Twitter. Here, we study the characteristics of users who author and vote in polls in the 2020 Query dataset (see Table 1). Since Twitter does not reveal the identity of poll voters, we consider poll retweeters and favoriters as likely voters. We study how the three user groups deviate from a representative sample of the U.S. population in terms of their age, gender (in comparison to the U.S. Census), and political ideology (in comparison to the 2020 U.S. presidential election exit polls). To quantify each of these user characteristics we apply state-of-the-art user attribute inference methods, described next.

Age and Gender Inference

Methods We employ the multilingual, multimodal, and multi-label machine learning tool *M3-Inference* under MIT license (Wang et al. 2019) to infer the gender and age of users. *M3-Inference* is a deep learning text and image model that uses usernames, profiles, and photos to infer age and gender with state-of-the-art accuracy while diminishing algorithmic bias in comparison to other approaches. Since the model additionally infers the likelihood that the given account represents an organization, we exclude from our analysis those users who exceed an organization score of 0.90.

Results First, we find that the fraction of males is about two times larger among poll authors than among exit poll

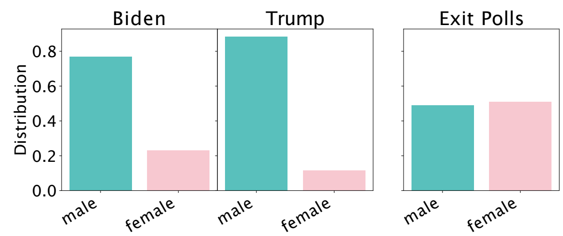

Figure 7: Gender distribution of authors of 2020 U.S. election polls on Twitter. For comparison, the rightmost figure shows gender distribution for the 2020 exit polls.

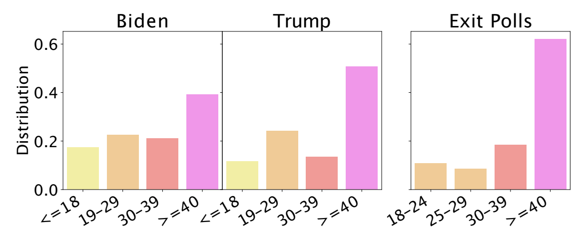

Figure 8: Age distribution of authors of 2020 U.S. election polls on Twitter. For comparison, the rightmost figure shows gender distribution for the 2020 exit polls. The bars are color-coded to mark the correspondence between the age brackets, e.g., the second bin for social polls corresponds to the first two age bins for exit polls.

respondents (see Figure 7) and that the fraction of poll authors below 30 years old is almost three times larger (see Figure 8). Our results conform to prior research suggesting that Twitter skews heavily male and young (Mellon and Prosser 2017). Polls won by Trump lean more male than those won by Biden (83 % vs 77.1%). These differences are greater among authors and their followers (58%) than among retweeters (53%) and favoriters (52%), suggesting that while young males are mobilized to create polls, people engaging with the polls are more similar to the general population in terms of age and gender. Retweeters and favoriters are comparably older (and more similar to U.S. voters) than authors and followers (42% and 50% have age $\geq$ 40 vs. 30% and 32%, respectively).

Political Ideology

Methods We estimate relative political ideology in Twitter polls using a Markov chain Monte-Carlo approach by Barberá et al. (2015); Barbéra (2015) under MIT license. The tool infers the political ideology of a account based on its followees. Specifically, if a user A follows a majority of right-wing accounts and user B follows a majority of left-wing accounts, the tool will output a positive value for user A and a negative value for user B. Each user instance is mapped to a continuous political ideology value in the interval $[-3, 3]$. While we provide the raw distribution of these values, in the next section (RQ2.5) we discretize this range into three bins (*Left*, *Moderate*, *Right*) for simplicity. Prior work shows that this approach performs comparably

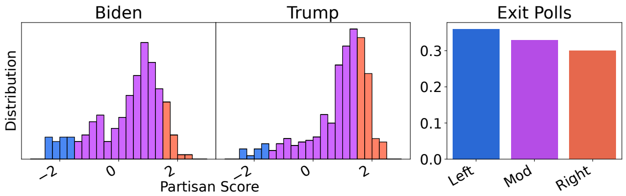

Figure 9: Distribution of partisanship scores for poll authors for social media polls, separately for Biden and Trump (left and center), and exit polls (right).

with standard ideological assessment surveys (Barberá et al. 2015). Due to rate limitations of the Twitter API and the removal of Academic API access, collecting the followees for Decahose users is infeasible.

Results To compare Twitter users' political ideology scores with the political affiliation data from exit polls, we convert the continuous scale of inferred political ideology of users into a discrete scale. Also, for our discussion, we use the terms political ideology and political affiliation interchangeably. Our analysis shows that the distributions of political ideology of poll authors is skewed towards the right (Figure 9). This result is consistent with our finding that social poll results are skewed towards Trump (RQ2.1). We also find that retweeters and favoriters of social polls are even more likely to be more conservative than the poll authors themselves (figure not shown). This asymmetry resembles the one observed in the distribution of political ideology of users who interact with misinformation, which also predominantly leans to the right (Nikolov, Flammini, and Menczer 2021; González-Bailón et al. 2023).

Correlates of Bias in Poll Outcomes (RQ2.5)

The previous sections paint a broad picture of the sources of bias that are potentially present in Twitter election polls. Next, we relate such biases to poll results, to assess the relative strength of different sources of bias.

Methods Here, we analyze polls from the 2020 query dataset. We perform ordinary regression to relate potential sources of bias to poll outcomes, operationalized as the fraction of votes cast in favor of Trump. Building on the observations from the previous sections, we consider potential sources of bias associated with the characteristics of both the authors of the polls and their respondents. Like in the previous section, we use retweeters and favoriters as proxies for poll respondents. In the regression model, we include as independent variables several sociodemographic attributes: the users' gender (male or *female*),[6] age group (*less than 30*, between 30 and 39, or greater than 40), political ideology (Democrat, *moderate*, or Republican), and location (U.S. red state, U.S. blue state, or *U.S. swing state*).[7] Furthermore, we include three additional user attributes: foreign location to the U.S. (yes or *no*), bot score (bot or *not bot*, using the threshold 0.83 for bot users), and hyperactivity (yes or *no*, using the threshold of 20 tweets/day for hyperactive users). Past work related such attributes to information operations, and as such they may relate to manipulation attempts.[8] We encode user traits as categorical for poll authors and as probability distributions of traits of retweeters or favoriters of a poll. To reduce noise in the dependent variable and the number of missing independent variables, we exclude polls with fewer than $M = 60$ votes. We choose M by maximizing R^2 and normalized Akaike information criterion (Cohen and Berchenko 2021). We impute missing values in the remaining polls ($N = 595$), substituting them with the mean of the corresponding independent variable.

Results The regression model explains a large fraction of the variance of poll outcomes, with an adjusted $R^2 = 0.514$. Thus, we further the analysis and interpret the coefficients associated with the independent variables. The model shows support for most of our hypotheses. Higher rates of support for Trump are expressed in polls with older demographics (the coefficients are statistically significant for retweeters), as well as users leaning Republican (authors and favorites). Conversely, higher support for Biden is associated with users leaning Democrat (retweeters and followers). We also find that users located in foreign countries (favoriters) are associated with polls that support Biden; this finding is in line with accounts of support for Biden and disfavor for Trump in NATO countries (Howorth 2021; Sintes-Olivella et al. 2022). Furthermore, the model shows a correlation between support for Trump and the presence of automated and hyperactive accounts among poll authors, which echoes previous findings about the heightened activity and strategic use of Twitter as a political medium by Trump supporters. Contrary to our intuition, we do not observe a connection between support for Trump and users' gender, despite a larger fraction of males in polls won by Trump (Figure 7); nor do we observe differences between users located in historically red or blue states: these associations are statistically insignificant when controlling the remaining variables in the model.

The quality of the model fit suggests that it may be possible to infer poll biases from the characteristics of the user base. However, the fact that significant coefficients in the model are associated with different, specific user groups asserts the importance of properly accounting for the different roles that users play in the generation of polls, their promotion, and engagement with them. Note that all independent variables are stable traits of the users that can be known in advance of the polls' outcomes, which makes them a promising avenue for predictive applications such as the post-stratification of poll outcomes to address their biases and the use of polls for the social sensing of public opinion.

[6] We indicate in italics the reference levels of the variables.

[7] We source the colors of the U.S. states from the list available on Wikipedia, based on the outcomes of the 2016 election: https://en.wikipedia.org/wiki/Red_states_and_blue_states

[8] We caution that such attributes do not necessarily imply malicious behavior. For example, political campaigns that schedule multiple tweets a day would likely be considered both bots and hyperactive, although their use of the platform is legitimate.

Independent variable	coef	P> |t|
const	0.42	***
p_a(gender=male)	0.01	
p_r(gender=male)	-0.01	
p_f(gender=male)	-0.03	
p_a(age$\in [30, 39]$)	-0.01	
p_r(age$\in [30, 39]$)	-0.01	
p_f(age$\in [30, 39]$)	0.00	
p_a(age≥ 40)	-0.01	
p_r(age≥ 40)	0.11	**
p_f(age≥ 40)	0.04	
p_a(ideology=dem)	0.00	
p_r(ideology=dem)	-0.15	*
p_f(ideology=dem)	-0.12	*
p_a(ideology=rep)	0.12	***
p_r(ideology=rep)	-0.03	
p_f(ideology=rep)	0.28	***
p_a(location=blue state)	0.03	
p_r(location=blue state)	-0.04	
p_f(location=blue state)	-0.03	
p_a(location=red state)	0.04	
p_r(location=red state)	-0.00	
p_f(location=red state)	-0.01	
p_a(location=foreign)	-0.03	
p_r(location=foreign)	-0.05	
p_f(location=foreign)	-0.09	**
p_a(bot=yes)	0.22	**
p_r(bot=yes)	0.01	
p_f(bot=yes)	0.02	
p_a(hyperactive=yes)	0.09	***
p_r(hyperactive=yes)	-0.08	
p_f(hyperactive=yes)	0.07	
Dependent variable:	% for Trump	
No. observations:	595	
Adj. R^2:	0.514	

Table 3: Coefficients for the ordinary least squares regression model using percent support for Trump as dependent variable. Coefficients with subscript a, r, and f refer to poll authors, retweeters, and favoriters, respectively. We indicate statistical significance at levels $p < 0.001$ (***), $p < 0.01$ (**), and $p < 0.05$ (*).

Polls Spreading Voter Fraud Beliefs (RQ3)

We identify polls that express skepticism towards traditional media's coverage of the elections, distrust poll results from mainstream sources, or question the legitimacy of the electoral process. We call such polls "conspiratorial," as they often express popular conspiracy theories about the elections. We estimate their overall number and the level of user engagement with the polls.

Methods To identify conspiratorial polls, we analyze the Decahose polls using election-related keywords. We take a human-centered approach as follows. First, two authors manually evaluated a random sample of 1,000 polls to code whether the question wordings or response options were conspiratorial. This initial coding round identified 19 conspiratorial polls (coder 1: N = 19, coder 2: N = 23) with high agreement between annotators (Krippendorff's α = 0.76). The two coders resolved disagreements through discussion with a third coder for the final annotation of this sample.

85 conspiratorial polls	
Retweets	541
Favorites	424
Votes	10,075
Followers	997,025

Table 4: The numbers of retweets, favorites, and followers of the identified 85 conspiratorial polls. The vote count is the total for the 47 (out of the 85) conspiratorial polls that remain available on Twitter.

We expanded the selection of conspiratorial polls by filtering Decahose polls using a dictionary of keywords typical of conspiratorial narratives, such as "voter suppression," "illegal votes," and "rigged" (the full list is in Appendix E). Two coders manually annotated all polls matching either keyword, which resulted in 66 additional conspiratorial polls, for a total of 85.

Results Knowing that there are about 19 conspiratorial polls among 1,000 random election polls, and building on the results of RQ1 that estimated a total of 130,000 election poll, we estimate that about 19*130=2,470 conspiratorial polls were posted on Twitter in 2020.

Thus, our set of 85 conspiratorial polls is a very small subset (about 3.4%) of the total—note that although the dictionary-based approach is straightforward, it might miss some conspiratorial polls. Nonetheless, the extended sample of conspiratorial polls sheds light on the reach of conspiratorial polls on social media (Table 4), their use, and the efforts to moderate such polls. Most of these conspiratorial polls were published after the presidential election day. Of the 85 identified conspiratorial polls, 38 were unavailable at the time of writing due to deletions (29 polls) and account suspensions (9 polls). This suggests that either poll authors removed their polls or there was a moderation effort on the platform to curb conspiratorial information.

Through qualitative coding, we unpack the themes in the content of the sample. Conspiratorial polls predominantly reflect skepticism toward the accuracy and neutrality of mainstream media polls. Many of them express a lack of trust in polls conducted by major news outlets and national polling organizations, with a recurring theme of challenging these polls' results through independent, user-conducted polls on social media platforms. There is a notable emphasis on achieving "unbiased" or "accurate" polls, indicating a perception that existing polls might be biased, especially toward Trump. The wording of the questions in some polls also brings up issues related to election fraud and voter suppression. This sample collectively underscores a mix of distrust in traditional polling methods, a desire for alternative, user-generated data, concerns about political bias and election integrity, and the contentiousness of the 2020 elections.

Discussion and Conclusions

According to our study, there were about 20 million votes cast in about 13,000 Twitter polls related to the 2020 presidential election. On average, these polls show 58% support for Trump and 42% for Biden, in a striking reversal of the true election result, in which Trump achieved 46.8% of the votes and Biden 51.3%. Moreover, we find that about 1.9% of polls on the 2020 elections questioned the validity of mainstream polls and/or election results. Future research can estimate the effect of such polls on the beliefs and voting intentions of people interacting with them.

Whenever platform transparency is lacking, researchers are forced to consider a breadth of potential explanations for otherwise opaque behavior. In the 2020 presidential election polls, we find signals of questionable votes, the activity of bots and hyperactive users, as well as demographic biases. In particular, we find that Twitter's poll system misreported public vote counts compared to what was shown to poll authors. In our experiments, these discrepancies neatly match purchased votes. This result and the opacity of Twitter that enabled the discrepancy are troubling. The difference in vote counts could be attributed to a bot classifier on Twitter's end, filtering out votes deemed suspicious. Considering that Twitter's business model incentivizes perceived activity and interest, completely filtering out suspicious votes may prove unprofitable. Under this framework, Twitter's interests as a platform run parallel to those of online astroturfers and, in effect, against that of the public good, which makes transparency efforts all the more necessary.

Our demographic analysis shifts the framing from *disinformation* to *misinformation*. The over-representation of certain population strata broadcasts a biased message, which contradicts and potentially breeds suspicion of institutional polling. We show that bias along age, gender, and partisan lines explains a large fraction of poll outcome variance.

The opacity of Twitter makes it challenging to estimate the – misinforming or disinforming – effect of such biases. Twitter's discontinuation of academic access programs exacerbates the problem. Data access to researchers ensured through policies such as the bipartisan Platform Accountability and Transparency Act proposed by U.S. legislators and the Digital Services Act in Europe, may be a first step towards a solution. However, we can also imagine action on Twitter's part (and other social media platforms, for that matter) to better inform its users about misleading biases in election polls.

Finally, we recognize that the non-representativity of Twitter users engaging in polls may be turned into useful information and leveraged as complementary to traditional representative surveys, and used in combination with them: hard-to-reach populations, such as those resistant to institutionalized polling, may be eager to disclose their preferences to ideologically aligned social poll authors.

Limitations

The findings of this study may not generalize beyond the polls gauging support for the 2016 and 2020 U.S. presidential election candidates. We look forward to similar studies for more recent elections, including elections in other countries, particularly the ones experiencing significant voter fraud conspiracies.

This study explains biases in poll outcomes with biases in the characteristics of users who might have participated in these polls. However, this relationship may be solely correlational, since (i) user demographics and political ideology are inferred after respective polls were conducted, (ii) polls showing biased outcomes may encourage votes from users having certain characteristics. Finally, further research is needed to establish a causal link between exposure to biased social poll outcomes and voter fraud beliefs.

Acknowledgments

This study was supported by the National Science Foundation award 2432051 and a Provost's Interdisciplinary Research Grant 2023 at University of Massachusetts Amherst. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Al-Rawi, A.; and Rahman, A. 2020. Manufacturing rage: The Russian Internet Research Agency's political astroturfing on social media. *First Monday*, 25(9).

Arnesen, S.; Johannesson, M. P.; Linde, J.; and Dahlberg, S. 2017. Do Polls Influence Opinions? Investigating Poll Feedback Loops Using the Novel Dynamic Response Feedback Experimental Procedure. *Social Science Computer Review*, 36(6): 735–743.

Barbéra, P. 2015. Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. *Political Analysis*, 23: 76–91.

Barberá, P.; Jost, J. T.; Nagler, J.; Tucker, J. A.; and Bonneau, R. 2015. Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber? *Psychological Science*, 26(10): 1531–1542. PMID: 26297377.

Benkler, Y.; Tilton, C.; Etling, B.; Roberts, H.; Clark, J. T.; Faris, R.; Kaiser, J.; and Schmitt, C. 2020. Mail-In Voter Fraud: Anatomy of a Disinformation Campaign. *SSRN Electronic Journal*.

Clifford, S.; and Sullivan, B. W. 2023. Do Survey Questions Spread Conspiracy Beliefs? *Journal of Experimental Political Science*, 10: 231–241.

Cohen, N.; and Berchenko, Y. 2021. Normalized information criteria and model selection in the presence of missing data. *Mathematics*, 9(19): 2474.

Das, M.; Toepoel, V.; and van Soest, A. 2011. Nonparametric Tests of Panel Conditioning and Attrition Bias in Panel Surveys. *Sociological Methods and Research*, 40: 32–56. Doi: 10.1177/0049124110390765.

Doherty, C.; Kiley, J.; Asheer, N.; and Jordan, C. 2020. Election 2020: Voters Are Highly Engaged, but Nearly Half Expect To Have Difficulties Voting. Technical report, Pew Research Center.

Ferrara, E.; Chang, H.; Chen, E.; Muric, G.; and Patel, J. 2020. Characterizing social media manipulation in the 2020 U.S. presidential election. *First Monday*, 25(11).

Ferrara, E.; Varol, O.; Davis, C.; Menczer, F.; and Flammini, A. 2016. The Rise of Social Bots. *Commun. ACM*, 59(7): 96–104.

FORCE11. 2020. The FAIR Data principles. https://force11.org/info/the-fair-data-principles/. Accessed: 2022-01-01.

Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Iii, H. D.; and Crawford, K. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92.

González-Bailón, S.; Lazer, D.; Barberá, P.; Zhang, M.; Allcott, H.; Brown, T.; Crespo-Tenorio, A.; Freelon, D.; Gentzkow, M.; Guess, A. M.; Iyengar, S.; Kim, Y. M.; Malhotra, N.; Moehler, D.; Nyhan, B.; Pan, J.; Rivera, C. V.; Settle, J.; Thorson, E.; Tromble, R.; Wilkins, A.; Wojcieszak, M.; de Jonge, C. K.; Franco, A.; Mason, W.; Stroud, N. J.; and Tucker, J. A. 2023. Asymmetric Ideological Segregation in Exposure to Political News on Facebook. *Science*, 381(6656): 392–398.

Harfoush, R. 2009. *Yes We Did! An inside look at how social media built the Obama brand*. New Riders.

Hickey, D.; Schmitz, M.; Fessler, D.; Smaldino, P.; Muric, G.; and Burghardt, K. 2023. Auditing Elon Musk's Impact on Hate Speech and Bots. arXiv:2304.04129.

Howorth, J. 2021. Europe and Biden: Towards a New Transatlantic Pact? *European View*, 20(1): 108–109.

Hughes, A. 2021. A Small Group of Prolific Users Account for a Majority of Political Tweets Sent by U.S. Adults. Technical report, Pew Research Center.

Hughes, A.; and Asheer, Ni. 2019. National Politics on Twitter: Small Share of U.S. Adults Produce Majority of Tweets. Technical report, Pew Research Center.

Hughes, S. G.-F.; Allbright-Hannah, K.; Goodstein, S.; Grove, S.; Zuckerberg, R.; Sladden, C.; and Bohnet, B. 2010. Obama and the power of social media and technology. *The European Business Review*, 16: 21.

Kalton, G.; and Schuman, H. 1982. The Effect of the Question on Survey Responses: A Review. *Journal of the Royal Statistical Society. Series A (General)*, 145(1): 42–73.

Keller, F. B.; Schoch, D.; Stier, S.; and Yang, J. 2020. Political Astroturfing on Twitter: How to Coordinate a Disinformation Campaign. *Political Communication*, 37(2): 256–280.

Lang, K.; and Lang, G. E. 1984. The Impact of Polls on Public Opinion. *The ANNALS of the American Academy of Political and Social Science*, 472(1): 129–142.

Marchlewska, M.; Cichocka, A.; and Kossowska, M. 2018. Addicted to answers: Need for cognitive closure and the endorsement of conspiracy beliefs. *European Journal of Social Psychology*, 48: 109–117.

Marsh, C. 1985. Back on the Bandwagon: The Effect of Opinion Polls on Public Opinion. *British Journal of Political Science*, 15(1): 51–74.

Mehta, C. 2022. Elon Musk's Twitter polls - from stake sales to CEO decisions.

Mellon, J.; and Prosser, C. 2017. Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Research & Politics*, 4(3): 2053168017720008.

Mislove, A.; Lehmann, S.; Ahn, Y.-Y.; Onnela, J.-P.; and Rosenquist, J. N. 2011. Understanding the Demographics of Twitter Users. *Int'l AAAI Conference on Weblogs and Social Media (ICWSM)*, 554–557.

Morwitz, V. G.; and Pluzinski, C. 1996. Do Polls Reflect Opinions or Do Opinions Reflect Polls? The Impact of Political Polling on Voters' Expectations, Preferences, and Behavior. *Journal of Consumer Research*, 23(1): 53–67.

NBC), N. E. P. A. N. C. C. 2020. National Election Pool Poll: 2020 National Election Day Exit Poll, 2020 [Dataset]. Roper #31119913, Version 4. Access Date: Oct-05-2023.

Nikolov, D.; Flammini, A.; and Menczer, F. 2021. Right and Left, Partisanship Predicts (Asymmetric) Vulnerability to Misinformation. *Harvard Kennedy School Misinformation Review*, 1(7): 1–13.

Pennycook, G.; Cannon, T. D.; and Rand, D. G. 2018. Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology*, 147: 1865–1880.

Pennycook, G.; and Rand, D. G. 2021. Research note: Examining false beliefs about voter fraud in the wake of the 2020 Presidential Election. *Harvard Kennedy School Misinformation Review*, 2(1).

Pfeffer, J.; Matter, D.; Jaidka, K.; Varol, O.; Mashhadi, A.; Lasser, J.; Assenmacher, D.; Wu, S.; Yang, D.; Brantner, C.; et al. 2023. Just another day on Twitter: a complete 24 hours of Twitter data. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 17, 1073–1081.

Price, V. E. 1992. *Public opinion*. Newbury Park: Sage. ISBN 080394022X 0803940238.

Sayyadiharikandeh, M.; Varol, O.; Yang, K.-C.; Flammini, A.; and Menczer, F. 2020. Detection of novel social bots by ensembles of specialized classifiers. In *Proceedings of the 29th ACM international conference on information & knowledge management*, 2725–2732.

Scarano, S.; Vasudevan, V.; Samory, M.; Yang, J.; and Grabowicz, P. 2024. Analyzing Support for U.S. Presidential Candidates in Twitter Polls. *Journal of Quantitative Description: Digital Media*, 4.

Schoch, D.; Keller, F. B.; Stier, S.; and Yang, J. 2022. Coordination patterns reveal online political astroturfing across the world. *Scientific Reports*, 12(1): 4572.

Sintes-Olivella, M.; Franch, P.; Yeste-Piquer, E.; and Zilles, K. 2022. Europe Abhors Donald Trump: The Opinion on the 2020 US presidential elections and their candidates in the European newspapers. *American Behavioral Scientist*, 66(1): 61–85.

Squire, P. 1988. Why the 1936 literary digest poll failed. *Public Opinion Quarterly*, 52(1): 125–133.

Suárez-Serrato, P.; Roberts, M. E.; Davis, C. A.; and Menczer, F. 2016. On the influence of social bots in online protests. Preliminary findings of a Mexican case study. *CoRR*, abs/1609.08239.

Uscinski, J. E.; Klofstad, C.; and Atkinson, M. D. 2016. What Drives Conspiratorial Beliefs? The Role of Informational Cues and Predispositions. *Political Research Quarterly*, 69: 57–71. Doi: 10.1177/1065912915621621.

Wang, Z.; Hale, S.; Adelani, D.; Grabowicz, P.; Hartman, T.; Flock, F.; and Jurgens, D. 2019. Demographic Inference and Representative Population Estimates from Multilingual Social Media Data. In *WWW '19: The World Wide Web Conference*, 2056–2067. ACM.

Wojcik, S.; and Hughes, A. 2019. Sizing Up Twitter Users. Pew Research Center. Https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/.

Zajaczkowski, G.; and Pereira, S. 2020. Strumień 600 botów. Ujawniamy, jak można kupić głosy w ankietach Twittera. (translation: Stream of 600 bots. We reveal how you can buy votes in Twitter Polls). TVP INFO. https://web.archive.org/web/20211027172609/https://www.tvp.info/48600557/sztuczne-wplywanie-na-wynik-ankiety-tak-dzialal-strumien-600-botow.

Zhang, Y.; Chen, F.; and Rohe, K. 2022. Social Media Public Opinion as Flocks in a Murmuration: Conceptualizing and Measuring Opinion Expression on Social Media. *Journal of Computer-Mediated Communication*, 27(1): zmab021.

Ethics Checklist

One crucial aspect of the design of this study is that it involved interacting with the authors of political polls through direct messaging. This enabled gaining unprecedented information on private vote counts, only visible to such users. After deliberation, the researchers decided not to disclose that the discrepancy may be the result of bought votes because the discrepancy was considered to be a phenomenon worth investigating regardless of this hypothesis. In particular, the poll authors may have colluded in the manipulation as well as have been in good faith. In either case and regardless of the truth of the hypothesis, the stigma associated with the presumption of collusion may have led to misreporting.

To gain evidence on the process of vote manipulation, the research design required purchasing specific numbers of votes by sellers of inauthentic user behavior. The authors took this matter seriously before determining its necessity for the study. The authors do not endorse such a market—in fact, the results of the present research aimed at exposing and ultimately reducing it, which was evaluated as a net positive when weighted against the small sum invested in the experiment. The authors took care of making the polls in the experiment invisible to regular Twitter users, so as not to expose the latter to manipulated results.

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes, because our paper releases no p.i.i. of the individuals in our dataset and has acted in accordance to IRB protocol.
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes, we clearly define our contributions and scope in the abstract and introduction.
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes, we substantiate machine learning outputs via (a) human coders as listed in *Classification and Validation of User Traits* and further emphasize the meaning of these results in RQ2.5. We note the relatively small number of respondents to our survey in RQ2.2. before drawing conclusions.
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? We do not generalize our dataset outside the context of Twitter election polling. Further, we allocate analysis and discussion to notable biases in the dataset itself when compared to election polling populations.
   (e) Did you describe the limitations of your work? Yes, we clearly mention it in the "API access" section, and address the limitations of manipulation identification in RQ2.2 and RQ2.3 as well as the conclusion.
   (f) Did you discuss any potential negative societal impacts of your work? While a breach of user privacy could have potential negative impacts, our work does not release and p.i.i of users in our dataset. We do not believe that any information provided could identify users.
   (g) Did you discuss any potential misuse of your work? Where appropriate, we note the size of samples before drawing conclusions (as in RQ2.2) and avoid take care to protect p.p.i of users in a paper with an explicitly political topic.
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? answerYesWe plan to provide all necessary data and scripts to replicate our findings, releasing only tweet ids to protect p.i.i.; releasing all collected information would leave profile pictures, self descriptions, and other potentially sensitive personal information exposed.
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes, we have read the ethics review guidelines and have ensured that our paper conforms to them.
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? Yes, we clearly outline the assumptions and also the possible limitations while calculating our results.
   (b) Have you provided justifications for all theoretical results? Yes, please check the Results and Discussion sections.
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? Yes, we discuss multiple competing hypotheses that might challenge our theoretical results. Check the hypotheses stated in the Introduction and our findings in the Results section.
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? Yes, we consider alternative mechanisms that could lead to getting the same results and discuss them in the Results and Discussion sections.
   (e) Did you address potential biases or limitations in your theoretical framework? We address the limitations of a smaller sample size in RQ2.2 and missing data in our dataset section.
   (f) Have you related your theoretical results to the existing literature in social science? Yes, we relate our methods and results to other previous work in the domain. Please check the Related Works section.
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? Yes, please check the Discussion section.
3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theoretical results? NA

(b) Did you include complete proofs of all theoretical results? NA

4. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? We provide a URL pointing to data and code necessary to replicate the findings of this paper.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? Yes. The goal here was to understand the attributes that contributed to the biases in the Twitter polls. Hence, the entire poll dataset was used to fit the linear regression model. The technique used to impute missing values is explained in Explaining Bias in Poll Outcomes under the methods section. There were no hyperparameters for the linear regression model. However, we ran multiple experiments with different variations to the features and chose the model that gave us the best AIC (Akaike Information Criterion) and R^2.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? Yes, in section addressing RQ2.5.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? NA
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? We describe our coding validation process for machine learning results in our Dataset section under *Classification and Validation of User Traits*
   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? We discuss our evaluation of the validity of classification both *Classification and Classification of User Traits* and RQ2.5 through regression analysis.

5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
   (a) If your work uses existing assets, did you cite the creators? We cite the creators of all three machine learning tools deployed for the study.
   (b) Did you mention the license of the assets? Yes, we mention the license of all code and models used.
   (c) Did you include any new assets in the supplemental material or as a URL? We provide a URL pointing to data and code necessary to replicate the findings of this paper.
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? While our university IRB does not require us to get consent from this data, the Twitter API provides no mechanism to ensure consent for data access. With respect to our survey experiment (RQ2.2), users voluntarily provided private vote counts at their discretion.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? We enforced strict access control mechanisms for the data. We do not attempt to track users across websites, we protect the anonymity of the users, we respect the context in which the content was shared, and we report results in aggregate.
   (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see FORCE11 (2020))? We describe the data release terms in the *Dataset* section.
   (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see Gebru et al. (2021))? We have not created a Datasheet for the dataset, but we have documented the data to ease replication.

6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
   (a) Did you include the full text of instructions given to participants and screenshots? Yes, we have provided the full message sent to participants in the Appendix.
   (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? Participation risk is not likely from our survey; however, we did mention our IRB approval.
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? Participants were not paid an hourly wage since they were merely contacted.
   (d) Did you discuss how data is stored, shared, and deidentified? The dataset has not be deidentified as it has not been released, but the paper analyses contain no p.i.i.

Appendix A

Table 5 contains the full message sent to poll authors.

Appendix B

The stores returned by Google are shown in Table 6.

Appendix C

To view the private vote count before December 2022, when Twitter change its user interface (UI), a user had navigate to "*View Tweet Activity*" from the tweet UI, then to "*View all engagements*" on the subsequent menu (Figure 10).

Appendix D

Table 7 contains the complete significance tests between political and random group distriubtions.

Appendix E

Full list of terms to match conspiratorial polls is provided in Table 8.

Hello! I'm a researcher at the University of Massachusetts, Amherst studying Twitter polls. I've noticed that in some instances on Twitter, the site's public votes shown on user timelines differ from the private votes listed in Tweet Analytics, and I'm studying potential causes and frequency of this phenomenon.
On **END_DATE** you posted **LINK_TO_TWEET**. Would you be willing to anonymously contribute to our study by disclosing the number of recorded votes displayed within Tweet Analytics? Your privacy will be fully preserved: we will study the numbers of votes on aggregate across hundreds of polls without revealing their individual values. If you choose to participate, we will share with you the findings of our study.
To participate in the study, please simply copy-paste the following number of votes as a response to this message. To see the number, please click on the above link and navigate to "View Tweet Analytics". There should be an attribute labeled "Votes" under the "View all engagements" button. Please simply respond to this message by copy-pasting that number (of "Votes" in "View Tweet Analytics"). Please double-check that you're copying the right number correctly.

Table 5: Message sent to surveyed authors regarding obtaining private vote counts

Rank	Name
1	viplikes.net
2	socialwick.com
3	buytwitterpollvotes.com
4	socialboss.org
5	gettwitterretweet.com
6	famousfollower.com
7	socialyup.com
8	rousesocial.com
9	instafollowers.com

Table 6: Top 10 Google ranking of "*buy Twitter poll votes*" query retrieved on 12/5/2021

Type	Distirubtions	p (MWU test)
Bot scores	Authors vs. Rand. Authors	$p < 0.001$
Bot scores	Retweeters vs. Rand. Rets.	0.001
Bot scores	Favoriters vs. Rand. Favs.	$p < 0.001$
Foreign	Authors vs. Rand. Authors	$p < 0.001$
Foreign	Retweeters vs. Rand. Rets.	$p < 0.001$
Foreign	Favoriters vs. Rand. Favs.	$p < 0.001$
Activity	Authors vs. Rand. Authors	$p < 0.001$
Activity	Retweeters vs. Rand. Rets.	$p < 0.001$
Activity	Favoriters vs. Rand. Favs.	$p < 0.001$

Table 7: Complete significance tests by user group

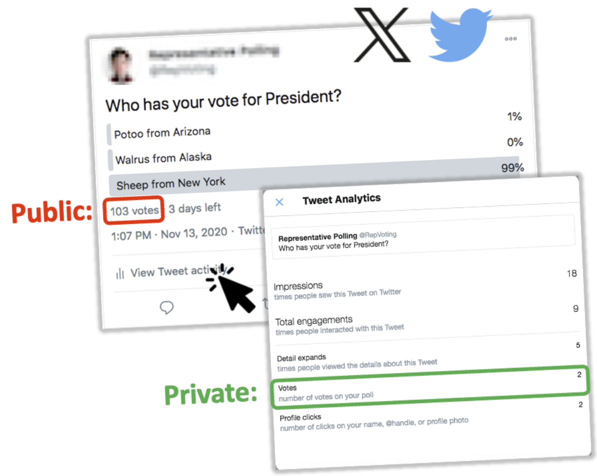

Figure 10: A poll used in the fake vote experiment. "Private" (green box) vote counts are distinct from those listed publicly (red box) under the tweet and are visible to the poll author under the "View Tweet Activity" tab. As a result of Twitter UI changes at the end of 2022, it is now substantially more complicated to see this count for new tweets and practically impossible for tweets from 2016.

#bidenlosestotrump, #votersuppression, accurate, accuracy, biased, blue wave crap, fraud, illegal ballots, illegal votes, illegal win, illegal winner, illegally win, illegally won, interesting statistic, legal ballots, legal votes, legal win, legal winner, legally win, legally won, real votes, real winner, recount, rejection, rejected, rigged, suppression, unbiased poll, voter suppression

Table 8: Full list of keywords used for identifying conspiratorial polls