Proceedings of the Twentieth International AAAI Conference on Web and Social Media (ICWSM 2026)

A Nonpartisan Study of Deepfake Activity and Engagement Around the 2024 US Presidential Election

Marco Postiglione[1], Isabel Gortner[1], Luke Fosdick[1], Chongyang Gao[1], Sarit Kraus[2], V.S. Subrahmanian[1]

[1]Northwestern University, Evanston, 60208, USA
[2]Bar-Ilan University, Ramat Gan, 5290002, Israel
{marco.postiglione, vss}@northwestern.edu

Abstract

We present the first quantitative study of deepfake activity during the 2024 U.S. presidential election by analyzing a novel dataset of 231 deepfakes (images, videos, and audio clips) across social media during the May-December 2024 election period. Our comprehensive statistical analysis examines five research questions: (1) whether deepfake publication spikes occur around key election events (KEEs), (2) whether there is a temporal relationship between deepfakes and KEEs (before, during, or after), (3) which specific types of KEEs trigger deepfake spikes, (4) whether KEEs boost engagement with deepfakes, and (5) the differential impact of various KEEs on deepfake engagement. Our findings reveal that spikes in deepfake activity preceded KEEs, and engagement with deepfakes (e.g., likes, comments) surged during pre-KEE time windows. Our curated dataset offers researchers a valuable resource to study the impact of synthetic media in political contexts, while our findings provide valuable advice for policymakers and social platforms to develop appropriate measures to counter potential malign deepfakes before future elections.

Introduction

Concerns about the use of deepfakes in elections have been growing. In Asia, deepfake videos of deceased Indonesian dictator Suharto were used to support a Presidential candidate in 2024[1]. In Africa, deepfakes have reportedly been involved in the 2023 Nigerian election (Emovwodo and Ayo-Obiremi 2024). Łabuz and Nehring (2024) study the use of deepfakes in 11 elections (many in 2023), including ones in Turkiye, Argentina, Poland, UK, France, Bulgaria, Taiwan, Indonesia, India, and Slovakia. These incidents have made synthetic and inauthentic media a top concern for election officials, both before and after the 2024 U.S. Presidential election[2]. *Though these past studies offer valuable insights, they lack a unified, systematically curated dataset of election deepfakes with comprehensive metadata suitable for rigorous statistical analysis. While fact-checking organizations may maintain collections of political deepfakes, these are typically dispersed across multiple sources and lack the standardized annotation necessary for systematic research.*

Figure 1: Examples (images and video frames) of deepfakes referenced in our USPED2024 dataset.

This paper is an attempt to address this gap by providing a highly curated dataset and a rigorous statistical study of deepfake use in and around the time of the 2024 U.S. Presidential election.

We investigate the relationship between *Key Election Events* (abbreviated KEEs) and deepfake activity in terms of release of deepfakes and the engagement garnered by such deepfakes. KEEs are defined as significant political developments during the election cycle that receive substantial coverage from reliable, non-partisan news sources and could potentially influence public discourse around the election. These events encompass various types of political developments including legal proceedings involving candidates, televised debates, party conventions, major candidacy announcements or withdrawals, security incidents, and high-profile campaign activities. By examining the temporal relationship between these events and deepfake publication patterns, we can better understand how political developments may influence the creation and dissemination of synthetic media content.

To achieve this, we first carefully curated a dataset of 231 deepfakes (169 images, 38 videos and 24 audios)[3]. Figure 1 shows a few representative examples. Each item was carefully annotated with relevant metadata such as the URL, the name of the candidate depicted, the content category, and engagement metrics (e.g., likes, comments, reposts). To sup-

[1]https://www.npr.org/2024/12/21/nx-s1-5220301/deepfakes-memes-artificial-intelligence-elections

[2]https://www.washingtonpost.com/technology/2024/11/09/ai-deepfakes-us-election/,https://abcnews.go.com/Politics/ai-deepfakes-top-concern-election-officials-voting-underway/story?id=114202574

[3]Videos with audio and visual components are counted independently as being both a video and an audio sample.

port future research, we make our dataset available to academics who agree to an ethical usage policy[4].

We then investigate five research questions using rigorous statistical methods:

- **RQ1**: Is there a steep increase (spike) in publication of deepfakes around key election events (KEEs)?
- **RQ2**: Do deepfakes tend to be published ahead of, during, or after KEEs?
- **RQ3**: Which specific types of KEEs (e.g. National Conventions, presidential debates, candidacy announcements) trigger spikes in deepfake publication activity?
- **RQ4**: Are KEEs associated with increased engagement (e.g. likes, comments, shares) with deepfakes on social media?
- **RQ5**: Are some KEEs associated with greater increases in engagement with deepfakes than others?

Our study provides empirical insights into these questions through the first systematic statistical analysis of deepfake activity during a U.S. presidential election. Each research question was operationalized through specific statistical hypotheses designed to test the core claims implied by the questions.[5]

Related Work

Recent research has called for more empirical studies to understand how AI-generated disinformation (Pei et al. 2024) operates in practice (Feuerriegel et al. 2023; Byman et al. 2023; Dalal et al. 2024; Wei, Xu, and Hui 2024). Our work responds to this call by providing a rigorous evaluation of how deepfake content correlates with KEEs during the U.S. 2024 presidential election and examining the engagement patterns that may amplify their reach. This section reviews existing literature relevant to our study, focusing on the use of deepfakes in elections, the temporal dynamics of misinformation, and the analysis of social media engagement with synthetic media.

Deepfakes in Political Contexts Research has shown that deepfakes of respected public figures can significantly enhance the believability of false information and may amplify the impact of disinformation campaigns by deceiving audiences and spreading misleading narratives (Ruffin et al. 2024). In the political sphere, studies have examined the potential for deepfakes to manipulate public opinion, disrupt elections, and erode trust in media and institutions (Vaccari and Chadwick 2020; Loewenstein 2024). Prior work has analyzed the use of manipulated media in past elections, identifying instances of both crude and sophisticated forgeries (Diakopoulos and Johnson 2021; Łabuz and Nehring 2024). The need for legislative and/or regulatory action has also been noted (Romero Moreno 2024)(Wei, Xu, and Hui 2024). This paper contributes to this growing body of work by providing a data-driven, statistical analysis of deepfake release during the 2024 U.S. Presidential Elections.

Timing and Human Perception of AI Disinformation The timing of a manipulated media release can greatly shape its potential consequences. Political events such as debates, conventions, and candidate announcements often act as high-salience moments when public attention peaks and attitudes may be more susceptible to influence (Sharma et al. 2019; Vosoughi, Roy, and Aral 2018; Gatta et al. 2023). Unlike traditional election misinformation that relies primarily on textual falsehoods or simple image edits, deepfakes represent a distinct threat category. Recent work on human perception of AI-generated misinformation has shown that audiences can find such content credible and persuasive, with effects on beliefs and attitudes (Spitale, Biller-Andorno, and Germani 2023; Bashardoust, Feuerriegel, and Shrestha 2024). These findings underscore the importance of mapping when deepfakes appear relative to KEEs, since the proximity of release to these events may heighten their potential to shape public discourse and voter perceptions.

Social Media Engagement with Synthetic Media High-profile political events that heavily impact communities have been shown to spur engagement on social platforms (Niles et al. 2019). Previous studies have explored how users engage with various forms of misinformation on social media, examining factors such as virality, user behavior, and platform algorithms (Shao et al. 2016; Vosoughi, Roy, and Aral 2018; Ceylan, Anderson, and Wood 2023). Research has also investigated engagement patterns associated with manipulated videos and images, highlighting the potential for these to garner significant attention and spread rapidly online (Wang et al. 2021). Our study extends this line of research by examining the temporal evolution of engagement metrics (likes, shares, comments, saves, views) and their relationship with KEEs, offering a nuanced, statistically validated understanding of how users interact with and potentially amplify deepfake content in a high-stakes political context.

The 2024 U.S. Presidential Election Deepfakes Dataset (USPED2024)

This study examines deepfake content disseminated during the 2024 U.S. Presidential Election cycle. Our data collection encompasses deepfakes published from May 1, 2024 to December 31, 2024, capturing the pre-election campaign period through post-election discourse. The dataset contains 231 deepfakes and includes 169 images, 38 videos and 24 audios. In addition, we identified and documented 21 Key Election Events or KEEs to analyze potential relations between deepfake proliferation patterns and significant election developments throughout this period. Although our analyses focus on events up to Election Day (November 5, 2024), we extended data collection through December 2024 to capture post-election deepfake activity aiming to support future research and provide the community with a more complete record of election-related deepfakes.

[4]This is identical to the ethical use approach proposed by others working on deepfake detection (Li et al. 2020; Dang et al. 2020). Link: https://sites.northwestern.edu/nsail/usped/.

[5]Please note that like most data-driven work, we study correlation, not causation, in these hypotheses.

Key Election Events (KEEs)

Key Election Events (KEEs) are politically significant developments during the 2024 U.S. Presidential election cycle that could potentially influence public discourse and synthetic media creation. To ensure objective event identification without partisan bias, we consulted the Ad Fontes Media Bias Chart[6] and selected three news sources rated as both *highly reliable* and *politically neutral*. Reliability scores range between 0 (contains inaccurate/fabricated info) and 64 (original fact reporting, high effort), while bias scores range from -42 (left) to 42 (right). Thus, reliability scores above 40 indicate fact-centered analysis, while bias scores between -6 and +6 indicate middle/centric/unbiased reporting. Selected sources are the BBC[7] (reliability score: 44.73; bias score: -1.33), Reuters[8] (reliability score: 45; bias score: -1.2), and the Associated Press (AP)[9] (reliability score: 44.82; bias score: -2.38). KEEs were included in our analysis if they received coverage from at least one of these neutral sources. We also included the Election Day (November 5, 2024) as a KEE due to its clear political significance.

Table 1 presents the resulting 21 KEEs selected from May 30, 2024, to the Election Day (November 5, 2024). These KEEs included legal proceedings, candidate debates, party conventions, campaign leadership changes, political rallies, and media appearances. For each event, we indicate which of the three news organizations provided coverage.

Deepfake Data Collection

To build a dataset of politically-relevant deepfakes, we sourced data from three channels: the Political Deepfakes Incidents Database (PDID) (Walker, Schiff, and Schiff 2024), accredited fact-checking organizations, and Google Alerts. PDID provides a continuously updated repository of politically salient deepfakes. To ensure broad coverage from May to December 2024, we supplemented PDID with deepfake content identified by established fact-checkers, including AFP Fact Check, the AI Incident Database, FactCheck.org, OpenSecrets, PolitiFact, and Snopes[10]. Additionally, we collected publicly available news articles, blog posts, and multimedia content flagged by custom Google Alerts configured with the keyword "deepfake" to collect deepfakes and filter those related to the 2024 U.S. Presidential Election.

All samples collected in this way ($N = 414$) then underwent a manual screening process. Specifically, each deepfake sample was evaluated to determine whether it depicted or referenced a political figure, such as a presidential candidate, elected official, news anchor, or politically associated public figure (e.g., Elon Musk). Samples that were not available online ($N = 59$), those sourced from websites rather than social media ($N = 16$), samples from social media platforms with too few deepfake entries ($N = 2$), and samples outside the designated May-December temporal window ($N = 106$) were excluded. This process resulted in a final dataset of $N = 231$ samples. During screening, we also addressed potential duplicates arising from cross-platform dissemination. When the same deepfake appeared on multiple platforms, it was recorded as a single sample in the dataset, with all known sources documented. Engagement data were attributed to the platform where the deepfake was accessible at the time of collection, ensuring that counts were not inflated by duplication.

For each selected sample, we recorded the following attributes: source, URL, publish date, source text/title, subject name, category, language, media type (video, audio, or image), online accessibility, number of likes, shares/reposts, comments, saves, views, misinformation warning flags (e.g., fact-check labels applied by platforms), original platform (e.g., X, YouTube), original collection source (PDID, fact-checkers, or Google Alerts), and political affiliation of the subject (Democrat, Republican, or both—e.g., deepfakes involving both Democratic and Republican candidates). It is important to note that the presence or absence of platform warning flags was not used as a selection criterion: rather, these flags were documented when present to provide a comprehensive record of platform interactions with the content for potential future research on content moderation effectiveness. A datasheet (Gebru et al. 2021) and codebook detailing the motivation, composition, collection process, recommended uses, and other relevant aspects of our USPED2024 dataset is included in the supplementary material.

An overview of the resulting dataset's content is presented in Figure 2. Figure 2(a) provides a visual representation of the key terms in deepfake posts around the 2024 U.S. Election cycle. Figure 2(b) illustrates the frequency of targeted political figures in the dataset, showing which political figures were referenced the most. Finally, Figure 2(c) shows the distribution of the targeted political parties across the dataset, highlighting that the Republican Party was targeted more than the Democratic Party. *We underscore that this study is designed to maintain analytical neutrality in reporting findings about deepfake targeting patterns. While our data may reveal asymmetric targeting of political figures, we are interested in studying the dissemination and engagement patterns of deepfakes in the context of the 2024 U.S. Presidential Election, without endorsing or criticizing any candidate, party, or topic.*

Results

RQ1: Deepfake Activity & KEEs

To address our first research question—whether there is a general relationship between the timing of deepfake releases and KEEs—we analyzed the daily number of newly published deepfakes from May 1, 2024 to December 31, 2024.

[6]https://adfontesmedia.com/

[7]https://www.bbc.com/news/videos/cdj39x21lxyo

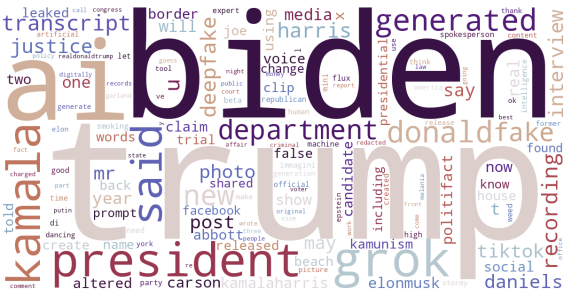
[8]https://www.reuters.com/world/us/key-dates-2024-us-presidential-race-2024-01-30/

[9]https://apnews.com/video/donald-trump-joe-biden-kamala-harris-chicago-pennsylvania-b51a77d48b594665b409d60a93de97d2

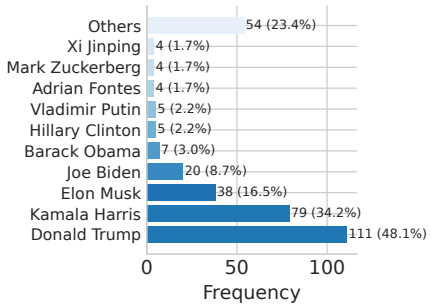
[10]Fact-checking websites: AFP Fact Check (https://factcheck.afp.com), AI Incident Database (https://incidentdatabase.ai), FactCheck.org (https://www.factcheck.org), OpenSecrets (https://www.opensecrets.org), PolitiFact (https://www.politifact.com), Snopes (https://www.snopes.com).

N.	Date	Description	BBC	Reuters	Associated Press
1	May 30, 2024	Trump found guilty on 34 counts of falsifying business records	✓		
2	Jun 27, 2024	Biden/Trump debate with concerns about Biden's fitness	✓	✓	✓
3	Jul 11, 2024	NATO Summit (Biden mistakes Zelenskyy for Putin)	✓	✓	
4	Jul 13, 2024	Attempted assassination of Trump	✓		✓
5	Jul 15, 2024	Republican National Convention begins	✓		
6	Jul 18, 2024	Trump accepts Republican nomination			✓
7	Jul 21, 2024	Biden withdraws from presidential race	✓		✓
8	Jul 24, 2024	Biden formally endorses Kamala Harris			✓
9	Jul 30, 2024	Harris's first campaign rally as presumptive nominee			✓
10	Aug 2, 2024	Harris announces candidacy	✓		
11	Aug 6, 2024	First Harris and Walz campaign rally	✓		
12	Aug 21, 2024	Democratic National Convention begins	✓		
13	Aug 22, 2024	Harris accepts Democratic nomination			✓
14	Sep 10, 2024	Presidential debate between Trump and Harris	✓		
15	Sep 15, 2024	Second assassination attempt on Trump in Florida	✓		
16	Sep 19, 2024	Harris speaks with Oprah at Michigan campaign rally	✓		
17	Oct 1, 2024	Vice Presidential debate	✓		
18	Oct 5, 2024	Elon Musk and Trump campaign rally	✓		
19	Oct 10, 2024	Obama speech at Harris campaign rally	✓		
20	Oct 16, 2024	Trump town hall with Univision and Harris Fox interview	✓		
21	Nov 5, 2024	Election day			

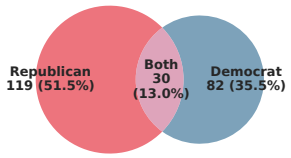
Table 1: Key Election Events (KEEs) related to the 2024 U.S. Presidential Election.



(a) Word cloud of key terms in deepfake posts around the 2024 U.S. Presidential Election cycle



(b) Frequency of targeted political figures



(c) Political affiliation distribution of deepfake targets

Figure 2: Overview of our 2024 U.S. Presidential Election Deepfakes Dataset (USPED), illustrating prevalent terminology used in deepfake discourse (a), the distribution of targeted political figures (b), and the partisan breakdown of subjects (c).

We identified significant spikes in this publication activity and tested for temporal correlations with KEEs.

Figure 3 shows the daily number of deepfakes published during the May-Dec 2024 time frame. The dashed vertical lines mark the dates of KEEs. The blue bars represent the daily deepfake count, while the green line shows the 7-day rolling average which provides a smoother trend. The shaded regions indicate periods of normal activity (light blue, within 2 standard deviations of the mean) and high activity (light red, more than two standard deviations above the mean). The descriptive statistics for the daily deepfake count (mean $\mu = 0.87$, standard deviation $\sigma = 4.35$, median = 0, max = 50) are also displayed in the upper left corner. High deepfake activity occurs on a day when over $\mu + 2\sigma = 9.6$ deepfakes are posted on that day.

Deepfake Spikes To identify statistically significant anomalies in deepfake publication activity, we calculated standardized z-scores for daily deepfake counts relative to the overall distribution mean and standard deviation. These z-scores were then converted to two-tailed p-values using the standard normal distribution approximation to test for significant deviations from expected daily publication rates. Six dates during August 2024 showed statistically significant spikes ($p < 0.05$) in deepfake publication activity. The most extreme outlier occurred on August 18, 2024, with 50 deepfakes ($z = 11.30$; $p = 10^{-28}$) published. Two additional major spikes were identified on August 15 and 17, each with 27 deepfakes ($z = 6.01$; $p = 1.9 \times 10^{-9}$) published. Three more dates showed moderate but significant spikes: August 14 (17 deepfakes; $z = 3.71$; $p = 2.1 \times 10^{-4}$), August 20 (15 deepfakes; $z = 3.25$; $p = 1.2 \times 10^{-3}$), and August 19 (13 deepfakes, $z = 2.79$; $p = 5.3 \times 10^{-3}$). This cluster of spikes spanning August 14-20 suggests the presence of triggering KEEs that substantially increased deep-

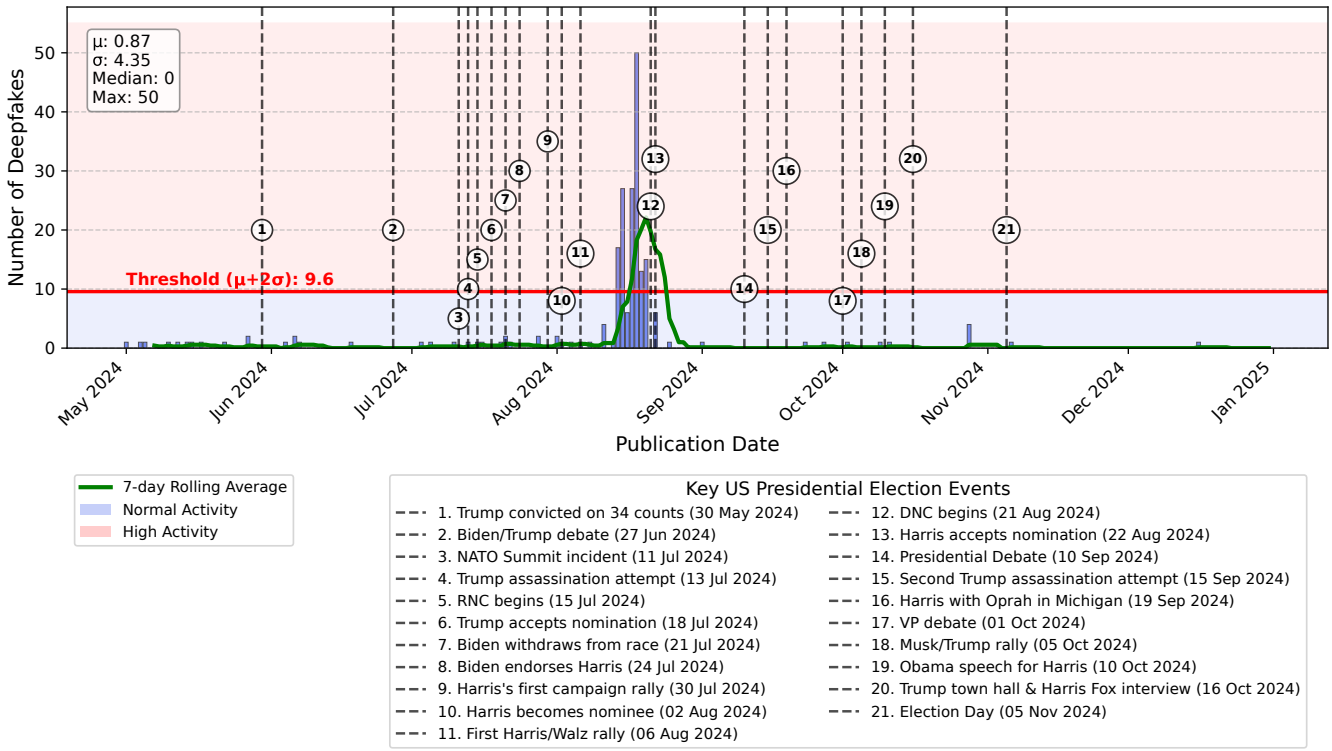


Figure 3: Daily count of deepfakes published around the 2024 U.S. Presidential Election in relation to Key Election Events (KEEs).

fake production during this week-long period.

Relationship Between Deepfake Publication Activity and KEEs To investigate the temporal relationship between deepfake activity and KEEs, we formulated and tested the following null hypothesis:

Hypothesis 1: *Key Election Events (KEEs) are no more likely to occur near periods of significant deepfake activity than during other periods.*

We operationalized this analysis using two key parameters. First, we defined "significant deepfake activity" using two thresholds: (1) days with ≥ 1 published deepfakes, representing any deepfake publication activity, and (2) days with $\geq \mu + 2\sigma = 9.6$ deepfakes, representing statistical spikes. Second, we applied two distinct temporal windows: (1) a "deepfake window" of size $w_d \in \{1, 3, 7\}$ days to identify dates within a distance of $\pm w_d$ days of significant deepfake activity, and (2) a "KEE window" of size $w_e \in \{1, 3, 7\}$ days to likewise determine whether dates fall near KEEs. For each of the 18 possible combinations of these parameters (2 thresholds $\times$ 3 deepfake window sizes $\times$ 3 KEE window sizes), we categorized all days in our study period as either within or outside a deepfake window, then calculated the *KEE rate* as the proportion of days in each category that were also within proximity of KEEs. We compared these proportions using the Mann-Whitney U test and applied the Benjamini-Hochberg false discovery rate (FDR) correction, which controls the expected proportion of false positives among the set of significant results, to account for multiple comparisons.

For example, consider the spike observed on August 18, 2024, with 50 deepfakes published. If we set $w_d = 3$, the corresponding deepfake window spans August 15 to August 21, capturing any KEE window that occurs within three days before or after the spike. For instance, the Democratic National Convention (DNC) began on August 21, falling exactly at the edge of this window. If we also apply a KEE window size of $w_e = 1$, then it would span from August 20 to August 22. The overlap between these two windows suggests temporal proximity between a significant spike in deepfake publication and a KEE.

Figure 4 shows a heatmap with the difference in KEE rates between days near and far from deepfake publication activity. Most heatmap cells are positive, indicating that KEEs tend to cluster around days with deepfake activity more than around other days.

For days with nonzero deepfake activity (threshold ≥ 1), we observed significant differences in 5 of 9 tested parameter combinations. The strongest relationship emerged with larger temporal windows, particularly when both the deepfake window and election event window were set to 7 days ($p = 10^{-7}$, FDR corrected). In this configuration, 70.4% of days near deepfake activity were also near KEEs, compared to only 21.4% of days without nearby deepfake activity. For days with deepfake activity spikes (threshold $\geq \mu + 2\sigma = 9.6$), 3 of 9 tested parameter combinations showed

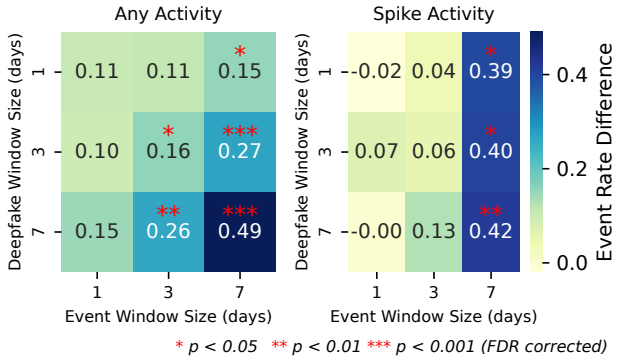


Figure 4: Heatmap showing the difference in Key Election Event (KEE) rates between days near and not near deepfake activity. Color intensity indicates the magnitude of difference, and asterisks (*) mark statistically significant results after FDR correction (* if $p < 0.05$, ** if $p < 0.01$, and *** if $p < 0.001$).

significant associations. When using a 7-day event window, all days (100%) with nearby deepfake spikes were also near KEEs, regardless of the deepfake window size. This perfect correspondence was statistically significant across multiple parameter combinations, with the strongest result occurring with a 7-day deepfake window and a 7-day event window ($p = 10^{-3}$, FDR corrected). The observed clustering patterns suggest opportunistic rather than coordinated behavior, lacking the synchronized activity signatures of known disinformation campaigns. This may reflect decentralized actors exploiting high-visibility events for maximum reach rather than orchestrated information operations.

Takeaway 1: ☞ *Deepfake activity significantly clustered around key election events (KEEs), especially during spike periods and with wider temporal windows, though patterns suggest opportunistic rather than coordinated behavior.*

RQ2: Temporal Alignment

The second research question examines the temporal alignment between deepfake publication activity and KEEs, to see if deepfakes tend to be published before, during or after KEEs. We analyzed the following hypothesis:

Hypothesis 2: *There is a positive correlation between KEEs and deepfake activity k days before or after each KEE.*

To quantify this, we computed the cross-correlation between a binary event series (1 = event day, 0 = non-event day) and the daily deepfake counts across various time offsets. In our analysis, a positive offset k measures the correlation between the event indicator on day t and the deepfake count on day $t + k$. Thus, positive offsets capture deepfake activity after KEEs, while negative offsets capture activity before KEEs. For instance, consider the Democratic National Convention that began on August 21, 2024: an offset of $k = -3$ would examine the number of deepfakes published three days before (August 18), potentially indicating anticipatory content creation. We evaluated these correlations over a two-week window spanning from $k = -7$ to $k = 7$ days to identify potential patterns in both directions.

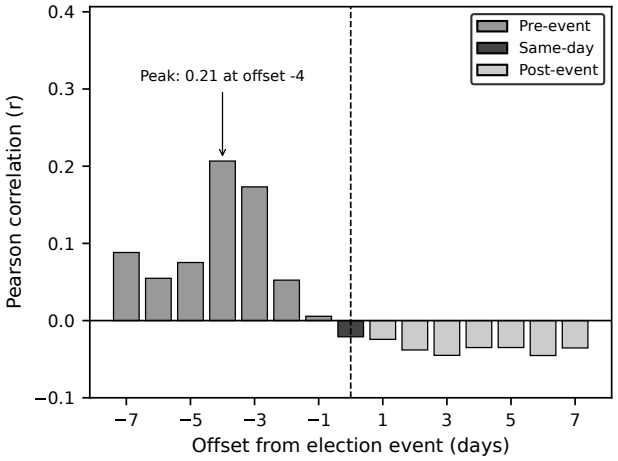


Figure 5: Cross-correlation between daily deepfake publication counts and Key Election Events (KEEs). Correlation coefficients are plotted for temporal offsets from seven days before each event (negative offsets) to seven days after (positive offsets).

Figure 5 shows the resulting Pearson correlation coefficients (r). Although the correlations are modest in magnitude, the strongest positive association occurs at $k = -4$ ($r = 0.21$), followed by $k = -3$ ($r = 0.17$). This suggests that major deepfake activity occurs 3-4 days before KEEs occur, rather than following them. Correlations at other pre-event offsets ($k \in [-7, -5] \cup [-2, -1]$) were considerably weaker ($r \in [0.001, 0.088]$). Interestingly, all post-event periods ($k \geq 0$) exhibited weak and negative correlation coefficients (ranging from $r = -0.02$ to $r = -0.05$), indicating a potential systematic decrease in deepfake production following KEEs.

Takeaway 2: ☞ *The strongest correlations observed between deepfake activity and KEEs occurred 3-4 days before the events, suggesting an anticipatory rather than reactive production pattern. But as these positive correlations are 0.17 and 0.21, this anticipatory pattern is generally not very strong.*

RQ3: Event-Specific Dynamics

Building upon the broader temporal patterns examined in RQ1 and RQ2, our third research question delves into event-specific relationships at a granular level. Given the finding that deepfakes tend to be published before KEEs (RQ2), we study temporal windows preceding KEEs. For example, we examine whether there was a significant increase in how many deepfakes were published in the 3 days preceding the

presidential debate on September 10, 2024, the 7 days preceding the general election on November 5, 2024, or the day preceding the NATO summit on July 11, 2024. To formalize this investigation, we tested the following null hypothesis:

Hypothesis 3: *The number of deepfakes published in the time period immediately before a specific KEE is not significantly different from the number of deepfakes published during the rest of the study period (May–December 2024) outside that time window.*

Our methodology involved three key steps: (1) defining temporal windows of 1, 3, and 7 days immediately before each event date, (2) categorizing each day in our study period as either falling within these target windows or outside them for each event, and (3) comparing the distribution of deepfake counts between these two categories using the non-parametric Mann-Whitney U test. To account for multiple comparisons across various event dates and temporal window sizes, we applied the Benjamini-Hochberg correction.

Figure 6 illustrates the mean deepfake counts within (blue bars) and outside (red bars) the 1, 3, and 7-day temporal windows preceding KEEs. Asterisks above the bars indicate statistically significant differences.

Our analysis shows that several KEEs were preceded by statistically significant increases in deepfake activity in the days leading up to the events. For example, the beginning of the Democratic National Convention on August 21 was preceded by a 7-day average of $\mu_7^{\text{in}} = 19.37$ deepfakes (August 15-21), compared to $\mu_7^{\text{out}} = 0.029$ outside this window, i.e. from May 1 to August 14 and from August 22 to December 31 ($p = 1.52 \times 10^{-55}$, FDR corrected). A similar pattern was observed for Democratic candidate Kamala Harris's nomination acceptance on August 22, with $\mu_7^{\text{in}} = 18.0$ versus $\mu_7^{\text{out}} = 0.034$ ($p = 1.52 \times 10^{-55}$, FDR corrected). Significant increases were also observed within shorter time windows for these and other events. On the day preceding the NATO Summit on July 11, the mean number of deepfakes was $\mu_1^{\text{in}} = 0.5$, compared to $\mu_1^{\text{out}} = 0.106$ outside the window ($p = 3.9 \times 10^{-5}$, FDR corrected). Similarly, the Democratic rally on August 6 saw an increase to $\mu_1^{\text{in}} = 0.625$, compared to $\mu_1^{\text{out}} = 0.104$ ($p = 2.82 \times 10^{-17}$, FDR corrected).

However, other KEEs, such as the September 10 presidential debate and the October 1 vice-presidential debate, did not show significant differences in deepfake counts within any of the tested temporal windows (all $p > 0.05$).

To examine whether deepfake targeting patterns differ by political affiliation around specific events, we conducted a complementary analysis comparing deepfake counts targeting Democratic versus Republican figures within the same temporal windows preceding KEEs. Deepfakes targeting both Democratic and Republican figures were not involved in this analysis. We tested the following null hypothesis:

Hypothesis 4: *The distribution of deepfakes targeting Democratic figures in the time periods immediately before KEEs is not significantly different from the distribution of deepfakes targeting Republican figures in the same time periods.*

	$\omega = 1$		$\omega = 3$		$\omega = 7$	
KEE Date	**Democrat**	**Republican**	**Democrat**	**Republican**	**Democrat**	**Republican**
May 30, 2024	0.00 ± 0.00	0.00 ± 0.00	0.25 ± 0.43	0.25 ± 0.43	0.12 ± 0.33	0.12 ± 0.33
Jun 27, 2024	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Jul 11, 2024	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.12 ± 0.33
Jul 13, 2024	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Jul 15, 2024	0.00 ± 0.00	0.50 ± 0.50	0.00 ± 0.00	0.25 ± 0.43	0.00 ± 0.00	0.12 ± 0.33
Jul 18, 2024	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.50 ± 0.50	0.00 ± 0.00	0.25 ± 0.43
Jul 21, 2024	0.00 ± 0.00	0.50 ± 0.50	0.00 ± 0.00	0.25 ± 0.43	0.00 ± 0.00	0.38 ± 0.48
Jul 24, 2024	0.00 ± 0.00	0.00 ± 0.00	0.25 ± 0.43	0.25 ± 0.43	0.12 ± 0.33	0.12 ± 0.33
Jul 30, 2024	0.00 ± 0.00	0.00 ± 0.00	0.25 ± 0.43	0.25 ± 0.43	0.12 ± 0.33	0.12 ± 0.33
Aug 2, 2024	1.00 ± 1.00	0.00 ± 0.00	0.50 ± 0.87	0.00 ± 0.00	0.38 ± 0.70	0.12 ± 0.33
Aug 6, 2024	0.50 ± 0.50	0.00 ± 0.00	0.25 ± 0.43	0.25 ± 0.43	0.38 ± 0.70	0.12 ± 0.33
Aug 21, 2024	2.50 ± 2.50	4.50 ± 4.50	9.25 ± 9.44	8.25 ± 6.98	7.50 ± 7.16	9.25 ± 7.74
Aug 22, 2024	0.00 ± 0.00	3.00 ± 3.00	3.00 ± 3.08	5.00 ± 3.24	6.75 ± 7.58	9.50 ± 7.60
Sep 10, 2024	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Sep 15, 2024	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Sep 19, 2024	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Oct 1, 2024	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.12 ± 0.33
Oct 5, 2024	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.25 ± 0.43	0.00 ± 0.00	0.12 ± 0.33
Oct 10, 2024	0.50 ± 0.50	0.00 ± 0.00	0.25 ± 0.43	0.00 ± 0.00	0.12 ± 0.33	0.00 ± 0.00
Oct 16, 2024	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.12 ± 0.33	0.12 ± 0.33
Nov 5, 2024	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00

Table 2: Mean daily deepfake counts targeting Democratic and Republican figures in temporal windows preceding Key Election Events (KEEs). Values shown as mean ± standard deviation. ω denotes the length of the temporal window in days.

Table 2 presents the mean daily deepfake counts (± standard deviation) targeting each party across temporal windows of $\omega = 1$, 3, and 7 days preceding each KEE. Around the most substantial deepfake activity, which occurred during the Democratic National Convention period (August 21-22, 2024), both parties were targeted, though with some variation in intensity across different temporal windows. For the 7-day window preceding August 21, Republican figures were targeted at a slightly higher rate ($\mu = 9.25 \pm 7.74$) compared to Democratic figures ($\mu = 7.50 \pm 7.16$). Similarly, for August 22 (Harris's nomination acceptance), Republican figures experienced higher targeting in the 7-day window ($\mu = 9.50 \pm 7.60$) compared to Democrats ($\mu = 6.75 \pm 7.58$). However, statistical testing using Mann-Whitney U tests revealed no significant differences in the distribution of deepfakes targeting Democratic versus Republican figures for any event within any temporal window (all $p > 0.05$ after Benjamini-Hochberg correction). The high variability in deepfake counts, as evidenced by the large standard deviations relative to the means, combined with the relatively small sample sizes within individual temporal windows, likely contributed to the lack of statistical significance.

Takeaway 3: ☞ *Deepfake activity spiked before party conventions and major rallies but not before debates, suggesting potential strategic targeting of specific political events. However, deepfakes targeted Democratic and Republican figures equally.*

RQ4: Social Engagement with Deepfakes

Following our analysis of deepfake publication patterns (RQ1, RQ2) and their event-specific relationships (RQ3), our fourth research question examines the dynamics of engagement with users on social media. We investigate how different engagement metrics evolve over time and whether specific KEEs correlate with higher engagement rates.

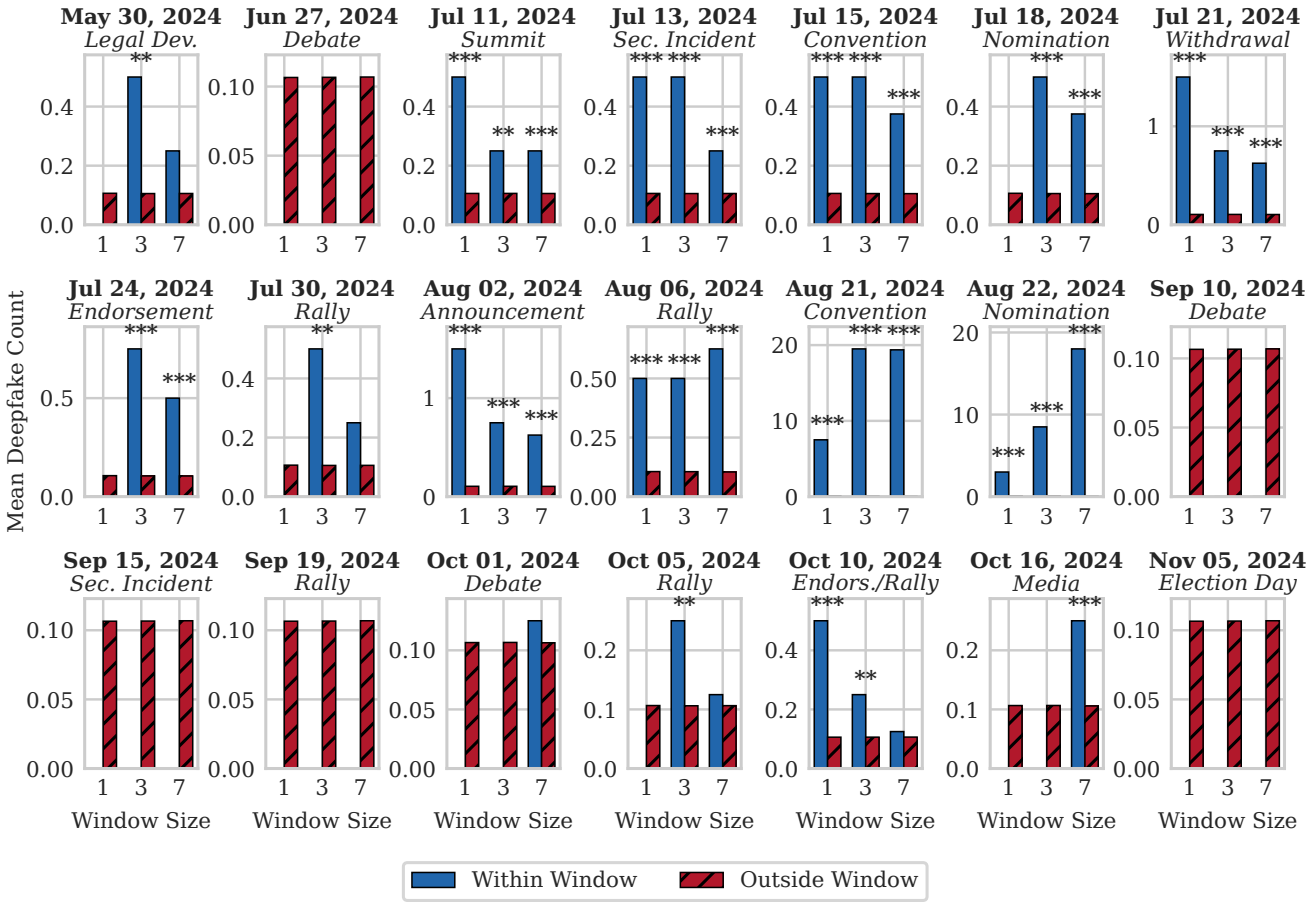


Figure 6: Mean deepfake counts within and outside temporal windows of 1, 3, and 7 days preceding Key Election Events (KEEs). Asterisks denote statistical significance based on Mann-Whitney U tests with Benjamini-Hochberg correction: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Note: y-axis scales vary across histograms.

Social Engagement Patterns We analyzed daily average engagement via five key metrics: likes, shares/reposts, comments, saves, and views. To enable comparison between these metrics with their inherently different scales, we normalized each engagement type by dividing daily values by the maximum value observed for that metric during the study period. This normalization produces relative engagement scores ranging from 0 to 1, allowing for direct comparison of temporal patterns across metrics.

Figure 7 shows normalized daily average engagement with deepfakes across all five metrics in relation to KEEs from May to December 2024. In most cases, engagement metrics move in tandem, suggesting that deepfakes that attract high engagement tend to do so across multiple metrics simultaneously. The engagement patterns show several notable spikes, particularly in July/August 2024, where there appears to be a significant peak on July 21 (Likes: 550,000; Shares/Reposts: 152,000; Comments: 28,630; Saves: 66,000; Views: 79,900,000).

Engagement & KEEs To examine whether engagement with deepfakes differs during periods preceding KEEs compared to non-event periods, we formulated and tested the following null hypothesis:

Hypothesis 5: *There is no significant difference in engagement metrics (likes, shares/reposts, comments, saves, and views) between temporal windows preceding key election events (KEEs) and the study period outside these windows.*

To test this hypothesis, we defined temporal windows of 1, 3, and 7 days (denoted ω) preceding each KEE and categorized deepfakes in our dataset based on their publication timing: those published within these pre-KEE windows (P_0) versus those published outside these windows (P_1). For each engagement metric, we compared the final engagement levels between these two groups of deepfakes using the Mann-Whitney U test and applied the Benjamini-Hochberg FDR correction to control for multiple comparisons. Prior to analysis, we excluded 45 samples with missing engagement data. These samples became unavailable between our initial data collection and engagement metric retrieval due to posts being removed by platform moderators, deleted by users, or

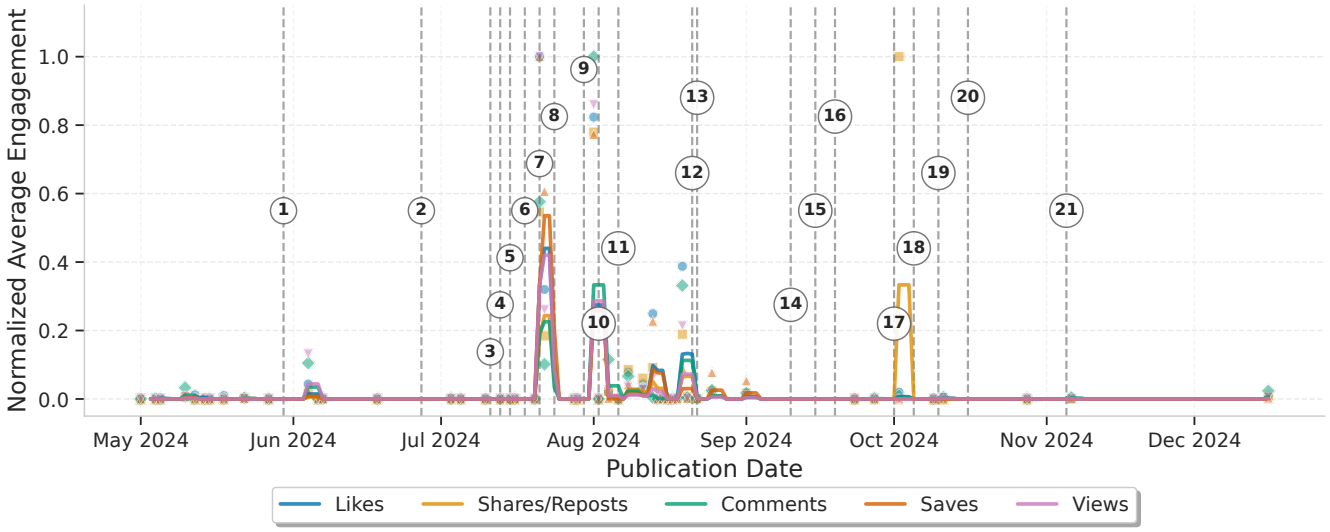


Figure 7: Normalized daily average engagement with deepfakes across five metrics (likes, shares/reposts, comments, saves, and views) from May to December 2024. The vertical dashed lines mark Key Election Events (KEEs), numbered consistently with Table 1. The 3-day rolling averages (solid lines) highlight general trends in engagement patterns.

	$\omega = 1$		$\omega = 3$		$\omega = 7$	
Metric	P_0	P_1	P_0	P_1	P_0	P_1
Likes	93,620	23,108	60,820	4,310	34,193	16,740
Shares/Reposts	19,018	4,386	11,796	886	6,609	3,620
Comments	4,189	978	2,675	141	1,503	541
Saves	10,721	1,029	4,158	429	2,331	1,755
Views	13,665,416	1,887,376	6,497,772	378,663	3,648,304	1,480,514

Table 3: Average engagement (likes, shares/reposts, comments, saves, views) for deepfakes published within temporal windows preceding KEEs (P_0) versus days outside these windows (P_1). Window sizes of 1, 3, and 7 days (ω) were analyzed for each metric.

made private.

Table 3 presents the mean values for each engagement metric within and outside the defined temporal windows. While engagement metrics are always higher preceding KEEs, our analysis revealed that only *Likes* demonstrate statistically significant differences when considering a 7-day window preceding KEEs ($p = 5.41 \times 10^{-4}$, FDR corrected). Specifically, within this window, deepfakes received an average of 34,193 likes compared to 16,740 outside.

Takeaway 4: ☞ *Deepfakes consistently drew more social media engagement (especially in terms of likes) before key election events (KEEs) as opposed to time periods that did not precede KEEs.*

RQ5: Event-Driven Social Engagement

While our previous analysis revealed some differences in engagement when considering all KEEs collectively (RQ4), we sought to determine if certain KEEs stood out as particularly influential. To investigate this question, we formulated the following null hypothesis:

Hypothesis 6: *There is no significant difference in engagement metrics (likes, shares/reposts, comments, saves, and views) between temporal windows preceding a specific key election event (KEE) and the study period outside these windows.*

We tested this hypothesis via a methodology similar to our previous analysis but focused on each event individually. For each of the KEEs in Table 1, we defined temporal windows of 1, 3, and 7 days preceding the event and compared engagement metrics within these windows to those in the broader study period. We used the Mann-Whitney U test for these comparisons and applied the Benjamini-Hochberg correction to control for multiple comparisons.

Despite the visually apparent spikes in engagement preceding certain KEEs in Figure 7, our analysis did not reveal any statistically significant differences for any individual event across all metrics and temporal window sizes after FDR correction, except for one case corresponding to the beginning of the Democratic National Convention on August 21. Interestingly, in the week leading up to this event, the average number of likes on published deepfakes was $\mu_7^{in} = 22{,}809$, compared to $\mu_7^{out} = 60{,}978$ outside this window.

To gain clearer insight into social media users' reactions to deepfakes, we extracted comments using Apify[11] and classified them into categories (Pro-Republican, Anti-Republican, Pro-Democrat, Anti-Democrat, Others) using LLaMA 3.2 Instruct (3B)[12]. Details on data retrieval and comment characteristics are provided in the Appendix. We tested the following hypothesis:

[11] https://apify.com

[12] https://huggingface.co/meta-llama/Llama-3.2-3B-Instruct

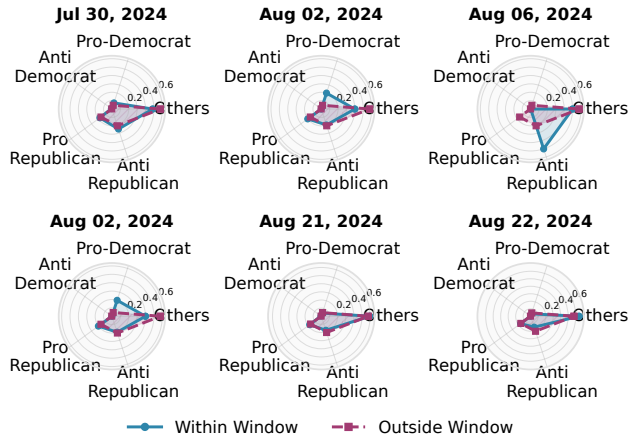


Figure 8: Party-linked comment distributions around KEEs in August 2024. Radar plots show the percentage of comments in each category (Pro-Republican, Anti-Republican, Pro-Democrat, Anti-Democrat, Others) for 3-day temporal windows preceding KEEs.

Hypothesis 7: *There is no significant difference in the likelihood of a comment belonging to a specific category (Pro-Republican, Anti-Republican, Pro-Democrat, Anti-Democrat, Others) between temporal windows preceding a specific key election event (KEE) and the study period outside these windows.*

For each KEE listed in Table 1, we defined temporal windows ω of 1, 3, and 7 days preceding each event. For each comment and category, we assigned binary indicators (1 if belonging to the category, 0 otherwise) and compared the distributions within temporal windows against baseline periods using Mann-Whitney U tests with Benjamini-Hochberg correction for multiple comparisons.

Figure 8 shows results for August 2024 events (i.e., the month coinciding with the deepfake activity spikes identified in RQ1) using $\omega = 3$ days. Each radar plot displays the percentage distribution of categories for comments occurring within versus outside the temporal window for a given event. Complete results across all time windows and KEEs are provided in the Appendix.

Most comparisons yielded non-significant differences due to sample size limitations and result variability. However, we observed statistically significant changes around Harris' acceptance of the Democratic nomination (August 22, 2024): Anti-Republican comments decreased from 21.03% to 15.28% within the temporal window (p =9.4e-04, FDR-corrected) while "Others" category comments increased from 57.77% to 64.58% (p =2e-03, FDR-corrected). As shown in the Appendix, a similar result is observed for the beginning of the DNC (August 21, 2024) with $\omega = 1$.

Takeaway 5: ☞ *Individual key election events (KEEs) generally did not trigger significant changes in deepfake engagement metrics, except for decreased likes before the Democratic National Convention. Party-related commenting activity showed significant changes before this KEE, with a decrease in Anti-Republican comments.*

Discussion

Our analysis reveals clear associations between deepfake publication activity and KEEs during the 2024 U.S. presidential elections. Our study of research questions RQ1-RQ3 found that deepfake activity clustered around KEEs rather than being randomly distributed. In RQ4 and RQ5, we examined social engagement metrics, finding that deepfakes published before KEEs generally attracted more engagement, though with important nuances across different KEEs.

In terms of RQ1, we found that days with significant deepfake activity were more likely to fall near KEEs, particularly when using larger temporal windows. This is consistent with existing research showing increased manipulation campaigns during critical political periods (Ferrara et al. 2020; Dobber et al. 2021), though they did not study deepfakes.

Our cross-correlation analysis in RQ2 revealed that deepfake activity peaked 3-4 days before KEEs rather than following them. This anticipatory pattern differs from traditional media coverage, which typically intensifies during and immediately after newsworthy events (Prior 2013). This finding extends previous observations that disinformation often precedes KEEs (Zilinsky et al. 2024; Starbird, DiResta, and DeButts 2023), suggesting a strategic deployment of deepfakes to prime public perception. The 3-4 day lead time may represent an optimal window for maximizing impact, allowing sufficient time for content to spread while remaining relevant to upcoming KEEs.

In RQ3, we observed substantial heterogeneity in deepfake responses to specific KEEs. For instance, there were huge spikes in deepfake activity preceding the Democratic National Convention and related events (August 21-22), with daily counts substantially higher than baseline levels in the preceding week. This concentration suggests that certain political events may be perceived as high-value targets for deepfake campaigns. The absence of increased deepfake activity before other high-profile events, such as presidential and vice-presidential debates, extends research by Badawy et al. (2019) by demonstrating that deepfake deployment appears selective and strategic rather than indiscriminate. These patterns suggest several possible strategic motivations, though we cannot definitively establish intent from observational data alone. First, the focus on high-profile events like conventions and rallies rather than targeted voter persuasion could represent efforts to undermine general trust in media and information rather than persuading specific voter segments. Second, the anticipatory timing patterns may indicate an agenda-setting effort designed to force campaigns and media to spend resources on defensive responses rather than preferred messaging. Third, the selective targeting of

certain KEEs over others might reflect a normalization strategy that gradually reduces public sensitivity to deepfakes in political discourse by making their presence routine around major events.

Our analysis of social engagement metrics in RQ4 showed that deepfakes published near KEEs generally attracted more engagement than those published during other periods, though this pattern was statistically significant only for likes within 7-day windows. This partially aligns with Pennycook and Rand's research (Pennycook and Rand 2021) showing emotionally charged or politically divisive content generates higher engagement. For RQ5, the absence of consistent, statistically significant engagement differences for individual KEEs suggests that engagement may be driven by content-specific factors rather than temporal context alone. A notable exception was the decreased number of likes preceding the Democratic National Convention, contradicting the general pattern of increased engagement before KEEs. This anomaly may reflect platform algorithmic interventions, coordinated reporting campaigns, or qualitative and content differences in deepfakes published during this period.

Our findings suggest that harmful effects of deepfakes during elections can be minimized through temporally-targeted approaches, with heightened vigilance 3-7 days before KEEs. These temporal patterns provide operational guidance for the framework proposed by Wei, Xu, and Hui (2024) for tackling AI disinformation, specifically suggesting that proactive monitoring and rapid response systems *before* KEEs may be more effective than post-hoc fact-checking. Effective proactive monitoring might combine temporal targeting with content-based signals, such as monitoring deepfakes of candidates likely to feature prominently in upcoming KEEs or tracking trending political topics that could become deepfake targets.

Limitations Our study has several limitations. First, despite combining multiple sources (PDID, fact-checking organizations, Google Alerts) and applying manual screening, platform or topic-specific biases may persist, and some deepfakes—especially in private networks or unflagged by fact-checkers—may remain undocumented. Nonetheless, the dataset is large enough for meaningful statistical analyses and provides a foundation for future research. Second, although our KEE identification relied on politically neutral sources (Ad Fontes rating –3 to +3), the heavier use of BBC coverage (Table 1) may introduce source-specific bias, even if the BBC is rated neutral. Third, while we find temporal correlations between KEEs and deepfake activity, these do not establish causality. Spikes could reflect heightened public attention, deliberate campaigns, platform algorithms, or coincidental engagement surges. Finally, our temporal windows (1, 3, 7 days) may not capture the full lifecycle of deepfake dissemination, including delayed or prolonged effects.

Conclusion & Future Work

This study offers the first statistical analysis of deepfake activity in the 2024 U.S. presidential election, revealing patterns in timing, dissemination, and engagement. While providing unprecedented insight into AI-driven disinformation, our findings also highlight the need for platform policies and detection systems attuned to event-driven surges. Preserving electoral integrity in the age of generative AI requires collaboration among researchers, platforms, and policymakers. Future work should extend to global elections to assess how cultural, geopolitical, and regulatory differences shape deepfake dynamics, and longitudinal studies could reveal how actors adapt across election cycles. Finally, we provide practical guidance for Regulatory Entities (REs)[13]. Because many KEEs are predictable, REs should intensify detection efforts in advance and maintain "surge teams" to respond rapidly to unexpected events. We recognize that implementation of these recommendations depends on voluntary cooperation between platforms and government agencies, as well as the development of appropriate information-sharing mechanisms that respect both platform autonomy and democratic oversight.

Ethical Statement

We followed strict ethical standards, using only aggregated data and excluding personally identifiable information from non-public individuals. Our study was reviewed by an Institutional Review Board (IRB)[14], which determined it does not constitute human subjects research. We also considered potential societal risks: while our findings reveal patterns in deepfake activity, we avoid tactical detail that could enable misuse and restrict dataset access through an ethical usage policy.

References

Badawy, A.; Addawood, A.; Lerman, K.; and Ferrara, E. 2019. Characterizing the 2016 Russian IRA influence campaign. *Social Network Analysis and Mining*, 9: 1–11.

Bashardoust, A.; Feuerriegel, S.; and Shrestha, Y. R. 2024. Comparing the Willingness to Share for Human-generated vs. AI-generated Fake News. *Proc. ACM Hum.-Comput. Interact.*, 8(CSCW2).

Byman, D. L.; Gao, C.; Meserole, C.; and Subrahmanian, V. 2023. *Deepfakes and international conflict*. Brookings Institution.

Ceylan, G.; Anderson, I. A.; and Wood, W. 2023. Sharing of misinformation is habitual, not just lazy or biased. *Proceedings of the National Academy of Sciences*, 120(4): e2216614120.

Dalal, A.; Gao, C.; Grimm, P. W.; Grossman, M. R.; Pulice, C.; Subrahmanian, V.; Tunheim, J.; and Linna Jr, D. W. 2024. Deepfakes in Court: How Judges Can Practively Manage Alleged AI-Generated Material in National Security Cases. *U. Chi. Legal F.*, 75.

[13]Regulatory Entities (REs) might include government entities that can issue advisories or guidelines (such as CISA), social media platforms' own content moderation teams, and third-party organizations (academic institutions, NGOs, fact-checking organizations) that monitor election integrity. While REs may lack direct regulatory authority over private platforms, they can coordinate information sharing, issue public advisories, and facilitate voluntary cooperation.

[14]Northwestern University, IRB ID: STU00223255.

Dang, H.; Liu, F.; Stehouwer, J.; Liu, X.; and Jain, A. K. 2020. On the Detection of Digital Face Manipulation. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, 5780–5789. Computer Vision Foundation / IEEE.

Diakopoulos, N.; and Johnson, D. 2021. Anticipating and addressing the ethical implications of deepfakes in the context of elections. *New media & society*, 23(7): 2072–2098.

Dobber, T.; Metoui, N.; Trilling, D.; Helberger, N.; and De Vreese, C. 2021. Do (microtargeted) deepfakes have real effects on political attitudes? *The International Journal of Press/Politics*, 26(1): 69–91.

Emovwodo, S. O.; and Ayo-Obiremi, I. 2024. The Implications of Deep Fakes Impact on Politics and Elections: The Nigerian Narrative. In *Navigating the World of Deepfake Technology*, 378–396. IGI Global.

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

Feuerriegel, S.; DiResta, R.; Goldstein, J. A.; Kumar, S.; Lorenz-Spreen, P.; Tomz, M.; and Pröllochs, N. 2023. Research can help to tackle AI-generated disinformation. *Nature Human Behaviour*, 7(11): 1818–1821.

Gatta, V. L.; Luceri, L.; Fabbri, F.; and Ferrara, E. 2023. The Interconnected Nature of Online Harm and Moderation: Investigating the Cross-Platform Spread of Harmful Content between YouTube and Twitter. In *Proceedings of the 34th ACM Conference on Hypertext and Social Media, HT 2023, Rome, Italy, September 4-8, 2023*, 39:1–39:10. ACM.

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

Łabuz, M.; and Nehring, C. 2024. On the way to deep fake democracy? Deep fakes in election campaigns in 2023. *European Political Science*, 23(4): 454–473.

Li, Y.; Yang, X.; Sun, P.; Qi, H.; and Lyu, S. 2020. Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, 3204–3213. Computer Vision Foundation / IEEE.

Loewenstein, S. 2024. Make America Fake again?: Banning Deepfakes of Federal Candidates in Political Advertisements under the First Amendment. *Fordham L. Rev.*, 93: 273.

Niles, M. T.; Emery, B. F.; Reagan, A. J.; Dodds, P. S.; and Danforth, C. M. 2019. Social media usage patterns during natural hazards. *PloS one*, 14(2): e0210484.

Pei, G.; Zhang, J.; Hu, M.; Zhai, G.; Wang, C.; Zhang, Z.; Yang, J.; Shen, C.; and Tao, D. 2024. Deepfake generation and detection: A benchmark and survey. *arXiv preprint arXiv:2403.17881*.

Pennycook, G.; and Rand, D. G. 2021. The psychology of fake news. *Trends in cognitive sciences*, 25(5): 388–402.

Prior, M. 2013. Media and political polarization. *Annual review of political science*, 16(1): 101–127.

Romero Moreno, F. 2024. Generative AI and deepfakes: a human rights approach to tackling harmful content. *International Review of Law, Computers & Technology*, 38(3): 297–326.

Ruffin, M.; Seo, H.; Xiong, A.; and Wang, G. 2024. Does It Matter Who Said It? Exploring the Impact of Deepfake-Enabled Profiles on User Perception towards Disinformation. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18, 1328–1341.

Shao, C.; Ciampaglia, G. L.; Flammini, A.; and Menczer, F. 2016. Hoaxy: A platform for tracking online misinformation. In *Proceedings of the 25th international conference companion on world wide web*, 745–750.

Sharma, K.; Qian, F.; Jiang, H.; Ruchansky, N.; Zhang, M.; and Liu, Y. 2019. Combating fake news: A survey on identification and mitigation techniques. *ACM transactions on intelligent systems and technology (TIST)*, 10(3): 1–42.

Spitale, G.; Biller-Andorno, N.; and Germani, F. 2023. AI model GPT-3 (dis)informs us better than humans. *Science Advances*, 9(26): eadh1850.

Starbird, K.; DiResta, R.; and DeButts, M. 2023. Influence and improvisation: Participatory disinformation during the 2020 US election. *Social Media+ Society*, 9(2): 20563051231177943.

Vaccari, C.; and Chadwick, A. 2020. Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social media+ society*, 6(1): 2056305120903408.

Vosoughi, S.; Roy, D.; and Aral, S. 2018. The spread of true and false news online. *science*, 359(6380): 1146–1151.

Walker, C. P.; Schiff, D. S.; and Schiff, K. J. 2024. Merging AI incidents research with political misinformation research: introducing the political Deepfakes incidents database. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 23053–23058.

Wang, Y.; Tahmasbi, F.; Blackburn, J.; Bradlyn, B.; De Cristofaro, E.; Magerman, D.; Zannettou, S.; and Stringhini, G. 2021. Understanding the use of fauxtography on social media. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, 776–786.

Wei, Z.; Xu, X.; and Hui, P. 2024. Digital Democracy at Crossroads: A Meta-Analysis of Web and AI Influence on Global Elections. In Chua, T.; Ngo, C.; Lee, R. K.; Kumar, R.; and Lauw, H. W., eds., *Companion Proceedings of the ACM on Web Conference 2024, WWW 2024, Singapore, Singapore, May 13-17, 2024*, 1126–1129. ACM.

Zilinsky, J.; Theocharis, Y.; Pradel, F.; Tulin, M.; de Vreese, C.; Aalberg, T.; Cardenal, A. S.; Corbu, N.; Esser, F.; Gehle, L.; et al. 2024. Justifying an invasion: When is disinformation successful? *Political Communication*, 41(6): 965–986.

Paper Checklist

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. Our study uses only publicly available, fact-checked deepfake data and politically neutral sources for analysis, ensuring privacy and avoiding bias. By making the fact-checked deepfake dataset and analysis methodology publicly available, it broadens access for researchers across institutions, regardless of their resources.

(b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes. Abstract and Introduction present the dataset and clearly state the research questions which our paper focuses on.

(c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes. Our methodological approach aligns tightly with each research question: we combine time-series visualization, spike detection (z-scores), Mann-Whitney U tests with FDR correction, and cross-correlation to analyze deepfake activity (RQ1–RQ3), and apply normalized metrics with non-parametric tests to assess engagement (RQ4–RQ5), all with justified, rigorously selected parameters and visualizations to support our claims.

(d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes. Our multi-source collection (PDID, fact-checkers, Google Alerts) and manual screening process mitigate artifacts from population-specific distributions, though we acknowledge residual platform- or topic-specific biases as noted in our Limitations.

(e) Did you describe the limitations of your work? Yes. A Limitations section is included after the Discussion of results.

(f) Did you discuss any potential negative societal impacts of your work? Yes. Our Ethical Statement outlines potential negative societal impacts.

(g) Did you discuss any potential misuse of your work? Yes. The Ethical Statement discusses potential misuse of our methodology, dataset and results.

(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? Yes. They are outlined in the Ethical Statement. Additionally, in the Deepfake Data Collection section, we specify that while we publicly release the dataset for research purposes, access requires agreement to an ethical usage policy that prohibits using the data to create new deepfakes or to harass depicted individuals. Our methodology for collecting and analyzing deepfakes is thoroughly documented to ensure reproducibility without requiring access to potentially harmful content.

(i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes. We ensured our paper is aligned with the Ethical Guidelines shared by the ICWSM 2026 Program Committee.

2. Additionally, if your study involves hypotheses testing...

(a) Did you clearly state the assumptions underlying all theoretical results? Yes. For each research question (RQ1-RQ5), we formally state the null hypotheses being tested and indicate non-parametric statistical methods, acknowledging that deepfake counts and engagement metrics may not follow normal distributions. Also, we explicitly state our use of the Benjamini-Hochberg correction to control for multiple statistical tests.

(b) Have you provided justifications for all theoretical results? Yes. After describing each result, we provide a comprehensive discussion in the Discussion section.

(c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? Yes. In the Limitations section, we discuss competing hypotheses that could challenge or complement our interpretation of deepfake surges during KEEs—for example, that such patterns may reflect increased public attention to election-related events or result from strategic communication by malign actors.

(d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? Yes. In the Limitations section, we consider alternative mechanisms such as algorithmic amplification, content recommendation dynamics, and unrelated fluctuations in user engagement, which could also explain the observed correlation between KEEs and deepfake activity.

(e) Did you address potential biases or limitations in your theoretical framework? Yes. Our theoretical framework was carefully designed to mitigate potential biases in data and events collection. With regard to data collection, we minimize selection bias by collecting data from three different sources (see *Deepfake Data Collection* section). To address potential political bias in event selection, we use the Ad Fontes Media Bias Chart to identify three highly reliable and politically neutral news sources (see *Key Election Events (KEEs)* section).

(f) Have you related your theoretical results to the existing literature in social science? Yes. Our study is grounded in and extends prior work across multiple domains of political communication, misinformation, and media effects. Our Related Work section provides an overview of the related literature. Our Discussion presents our findings in relation to existing literature.

(g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? Yes. Our Discussion section provides insights on both policy and practical implications for social media platforms. Additionally, our Conclusion outlines future research opportunities. The last paragraph also suggests regulatory policies.

3. Additionally, if you are including theoretical proofs...

(a) Did you state the full set of assumptions of all theoretical results? N.A.

(b) Did you include complete proofs of all theoretical results? N.A.

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)? N.A.

(b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? N.A.

(c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? N.A.

(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)? N.A.

(e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? N.A.

(f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? N.A.

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? Yes. Part of our dataset has been gathered from the Political Deepfakes Incidents Database (PDID) (Walker, Schiff, and Schiff 2024). We appropriately cited the creators.

(b) Did you mention the license of the assets? N.A. — no license is associated with the PDID dataset.

(c) Did you include any new assets in the supplemental material or as a URL? Yes. We provide the dataset used in our study as supplementary material to guarantee a transparent review process. For the sake of anonymity, we did not include a URL to the dataset as yet in the paper, but we will include it in the camera-ready version. Access to the dataset will be granted only to individuals from academic institutions who agree to an ethical usage policy.

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? Yes. In our study, we collected data involving well-known public figures. The Institutional Review Board (IRB) at the university where most of the authors are affiliated reviewed our data collection procedures and determined that the activity did not require signed consent forms. We specify this in the Ethical Statement.

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes. Given its nature, deepfakes in our dataset may contain offensive content targeting public figures. To avoid helping the dissemination of these deepfakes, we restrict access to our dataset only to individuals from academic institutions who agree to an ethical usage policy.

(f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see **?**)? Yes. A summary of how we intend to make our dataset FAIR is provided as follows:

- *Findable (F).* We have registered our dataset on Zenodo, a safe and trusted repository which assigns a Digital Object Identifier (DOI) to each uploaded resource. Data is associated with rich metadata and each dataset sample is properly identified. To minimize potential misuse of our work, we restrict the access to individuals from academic institutions who agree to an ethical usage policy.
- *Accessible (A).* Zenodo guarantees that data is permanently accessible and retrievable to all the emails to which access has been granted.
- *Interoperable (I).* Data is shared with a CSV format.
- *Re-usable (R).* Data is released with a Creative Commons Attribution-NonCommercial 4.0 (CC BY-NC 4.0) license, which allows others to use, share, and adapt our data for non-commercial purposes only, with attribution.

(g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see Gebru et al. (2021))? Yes. The datasheet is included in our Zotero repository and has been attached as supplementary material for review purposes.

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? N.A.

(b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? We collected data involving well-known public figures. The Institutional Review Board (IRB) at the university where most of the authors are affiliated reviewed our data collection procedures and determined that the activity did not require signed consent forms. We specify this in the Ethical Statement.

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? N.A.

(d) Did you discuss how data is stored, shared, and deidentified? N.A.

A Social Media Comments

Social media comments were collected from X (formerly Twitter) and Instagram posts using the Apify[15] platform, resulting in an initial dataset of 6,720 comments across both platforms. After data cleaning, the dataset was reduced to 5,577 comments. Topic modeling was performed using BERTopic, which leverages transformer-based embeddings rather than traditional TF-IDF vectors to capture semantic similarity between comments before clustering them into

[15]https://apify.com

thematic groups. The BERTopic pipeline generates high-dimensional semantic embeddings for each comment, applies dimensionality reduction, and uses density-based clustering to identify coherent topics.

Figure 9 presents a t-SNE visualization of the document embedding space, where each point represents a comment and colors indicate topic assignments. The visualization reveals the natural clustering structure discovered by BERTopic, with the ten largest topic clusters (Topics 1–10) highlighted using distinct markers to emphasize their prominence in the dataset.

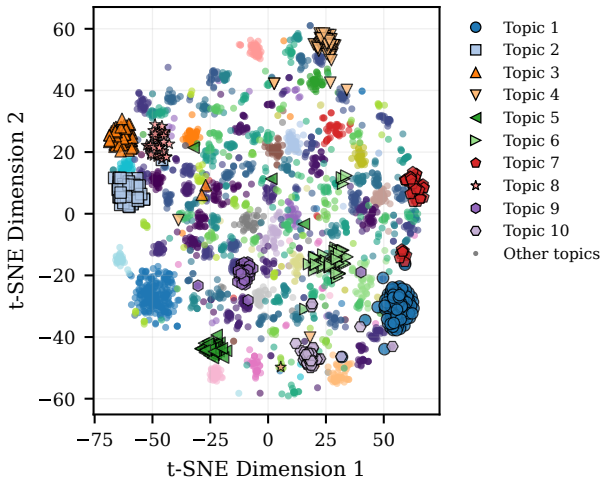


Figure 9: t-SNE visualization of BERTopic embeddings showing the distribution of social media comments in two-dimensional embedding space. Each point represents a comment, with colors indicating topic assignments. The ten largest topic clusters (Topics 1–10) are highlighted with distinct markers and included in the legend.

For interpretability, the ten largest clusters were selected and their comments were processed using the Llama-3.2-3B-Instruct language model[16] to generate single-sentence summaries capturing the central theme of each topic cluster. We used the prompt shown as follows:

Prompt

```
Summarize the following comments in one
sentence:

Comments: {topic_comments}

Summary:
```

Table 4 shows the 10 topic clusters ranked by size, from the largest cluster (192 comments about mixed reactions) to the smallest (54 comments about enthusiastic anticipation). The summaries capture the key themes ranging from political discussions about various figures to reactions to deepfake content and general social media engagement patterns

[16]https://huggingface.co/meta-llama/Llama-3.2-3B-Instruct

Comment Party-linked Categories Llama-3.2-3B-Instruct has also been used to extract categories (Pro-Republican, Anti-Republican, Pro-Democrat, Anti-Democrat, Others) to support our analysis in Hypothesis 7. The prompt is shown as follows:

Prompt

```
Assign a category (Pro-Republican, Anti-
Republican, Pro-Democrat, Anti-Democrat,
Others) to the following comment:

Comment: {comment}

Category:
```

This classification yielded 1,026 Anti-Republican, 948 Pro-Republican, 239 Pro-Democrat, and 29 Anti-Democrat and 3,335 ”Others” comments, revealing a dataset dominated by non-partisan content and a notable asymmetry between Republican-focused (1,974 total) and Democrat-focused (268 total) content.

Comments Analysis (full results) We now present the complete results of our comment analysis around key election events (KEEs). While the main text focused on August 2024 events using 3-day temporal windows (Figure 8), Figures 10, 11 and 12 provide comprehensive results across all KEEs listed in Table 1 and all three temporal window sizes ($\omega = 1, 3, 7$ days).

As noted in the main text, most comparisons yielded non-significant differences due to sample size limitations and result variability. However, several statistically significant changes were observed, summarized in Table 5. The most notable patterns include: (1) decreased Anti-Republican comments and increased ”Others” comments around July 24, 2024 (Biden’s withdrawal announcement) across multiple temporal windows, (2) increased Pro-Democrat comments around mid-July 2024 events (July 13-18, corresponding to the Republican National Convention period), and (3) the previously highlighted changes around the Democratic National Convention in August 2024.

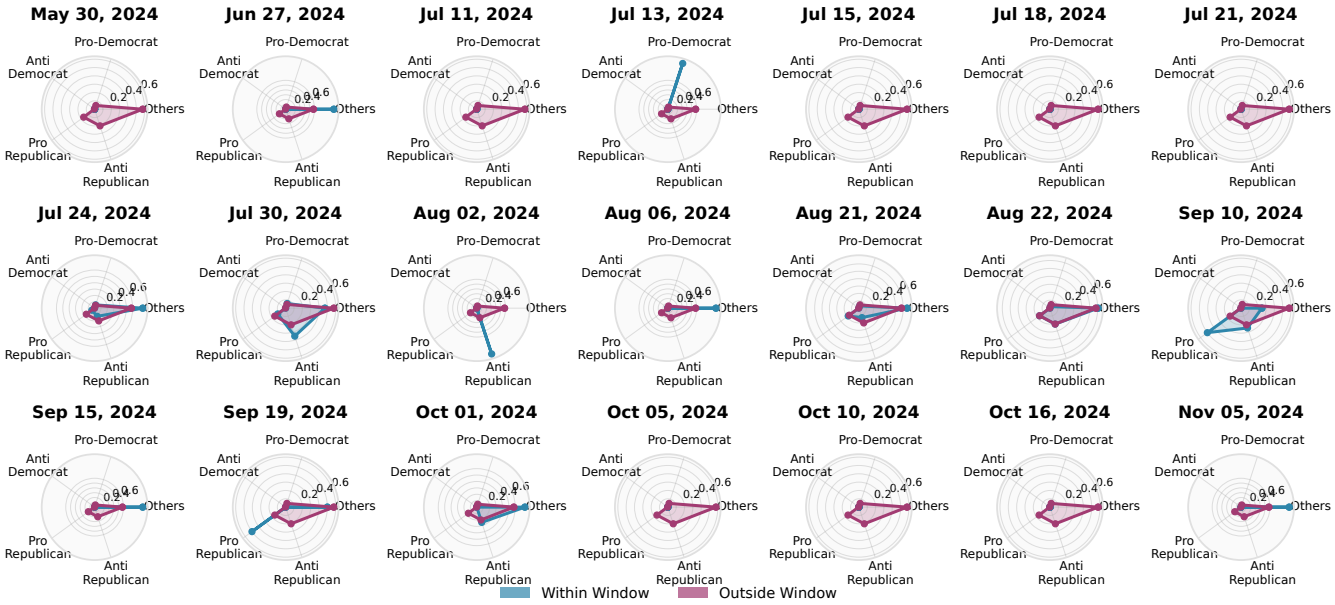


Figure 10: Party-linked comment distributions around all key election events (KEEs) using 1-day temporal windows.

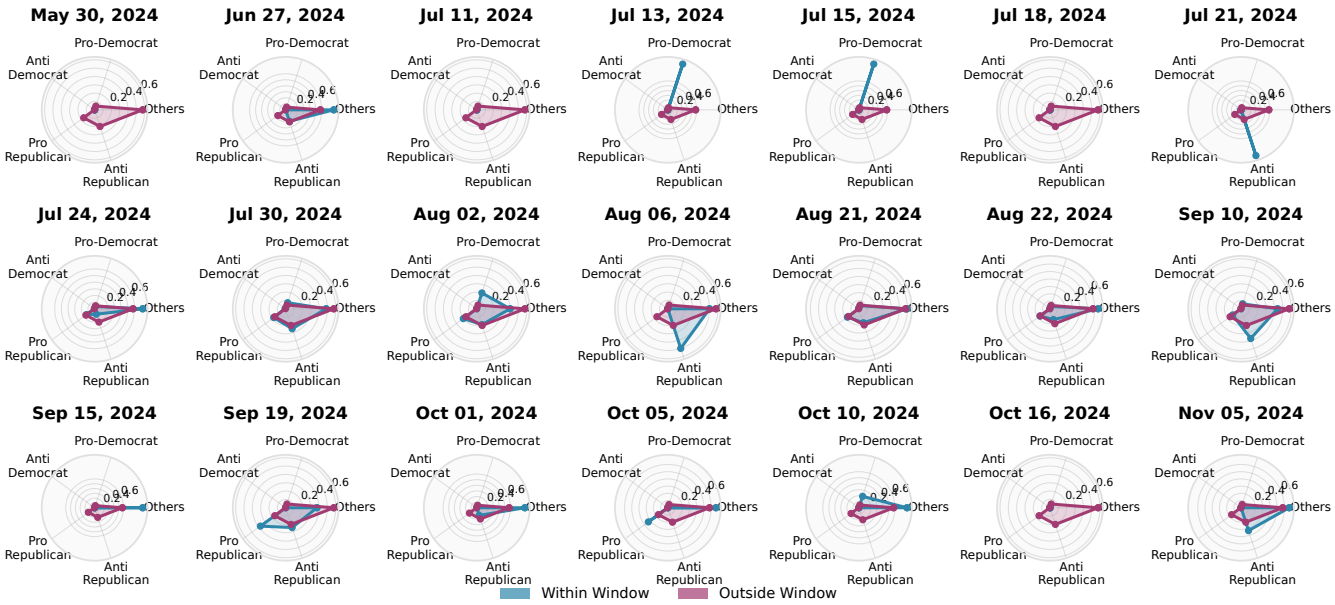


Figure 11: Party-linked comment distributions around all key election events (KEEs) using 3-day temporal windows.

Topic	Summary	n
1	The comments are a mix of enthusiastic and sarcastic reactions to something, ranging from ”I love it!” to ”What a joke!” and ”That’s perfect!”. There are also some comments that are simply ”yes” or ”no” or ”omg”, and some that are just emojis or phrases like ”Lmao” or ”Venom”. Overall, the comments are quite varied and playful, with a tone that is generally lighthearted and humorous.	192
2	Commenters express their dislike for Kamala Harris, calling her dumb, a failure, and a ”commi” (communist), and questioning her qualifications and American identity. Many also mock her, using humor and sarcasm to criticize her policies and actions. Some commenters praise Donald Trump, comparing him to Kamala Harris and suggesting that Trump’s popularity is due to his ability to ”destroy” her. Others express support for Trump and criticize Kamala Harris, calling her ”fake” and ”fraudulent.” Overall, the comments are overwhelmingly negative and hostile towards Kamala Harris.	134
3	Users are responding to a post by Elon Musk on Twitter, with some expressing support and others criticizing him for his views, while some are joking about his potential candidacy for president or his fashion sense. Note: The comments are a mix of serious and sarcastic responses, and some of them may not be suitable for all audiences.	114
4	Commenters argue about whether Joe Biden, Kamala Harris, or other Democrats are communists, with some claiming they are spreading misinformation and others claiming that communism is being misrepresented or misunderstood. Some commenters defend communism and others attack it, with many using inflammatory language and personal attacks.	113
5	Many commenters are excitedly discussing the idea of a new product, likely a flavor of chips, that will come in two colors: blue and red, with the blue color being associated with the Democratic Party. Some commenters are making humorous and lighthearted comments, while others are expressing their enthusiasm for the blue color. Several commenters mention the idea of the blue color being associated with a ”blue wave” or a blue wave election, and some make jokes about the idea of the red color being associated with the Republican Party. Overall, the comments are playful and celebratory.	76
6	The comments are overwhelmingly negative, with many users expressing racist, sexist, and homophobic slurs, making fun of the woman’s appearance, and questioning her credibility and authenticity. Some users also make baseless accusations against the woman, such as being drunk or having a personal relationship with someone else. The comments are often sarcastic and mocking, with some users pretending to be supportive or complimentary, but actually being cruel and dismissive.	71
7	Users are reacting to a deepfake video of Joe Biden, in which he appears to be saying ”Biden out,” and are responding with a mix of shock, amusement, and frustration, with some even joking about his age, health, and abilities.	67
8	The comments on the post are overwhelmingly negative, with many users expressing concern about the use of deep fakes, labeling the video as a ”deep fake,” and warning others about the potential dangers of deep fakes, while others make light of the situation with humor and sarcasm.	64
9	These comments express a mix of support and opposition to the potential female presidential candidate, with many users expressing skepticism, anger, and conspiracy theories about her and her campaign, while others express support and admiration for her.	62
10	The comments are overwhelmingly enthusiastic and optimistic, with many users expressing excitement and anticipation for a specific event or announcement.	54

Table 4: Top 10 Topic Clusters from Social Media Comments Analysis.

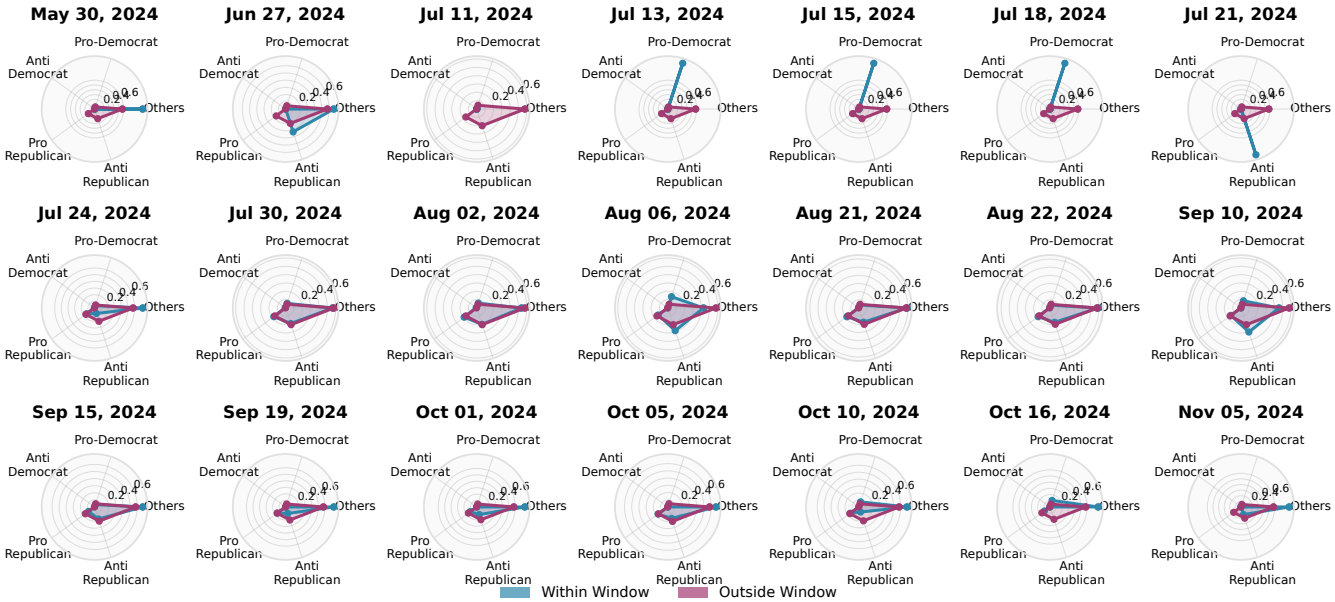


Figure 12: Party-linked comment distributions around all key election events (KEEs) using 7-day temporal windows.

Event Date	Window (days)	Category	Within Window	Outside Window	p-value	p-corrected
2024-07-24	7	Anti-Republican	0.087	0.210	5.60e-12	7.00e-10
2024-07-24	3	Anti-Republican	0.086	0.210	3.08e-12	7.00e-10
2024-07-24	3	Others	0.724	0.577	1.08e-10	9.01e-09
2024-07-24	7	Others	0.722	0.577	1.49e-10	9.34e-09
2024-08-21	1	Anti-Republican	0.135	0.210	2.66e-06	1.33e-04
2024-07-13	3	Pro-Democrat	1.000	0.047	8.06e-06	1.83e-04
2024-07-15	3	Pro-Democrat	1.000	0.047	8.06e-06	1.83e-04
2024-07-13	1	Pro-Democrat	1.000	0.047	8.06e-06	1.83e-04
2024-07-15	7	Pro-Democrat	1.000	0.047	8.06e-06	1.83e-04
2024-07-18	7	Pro-Democrat	1.000	0.047	8.06e-06	1.83e-04
2024-07-13	7	Pro-Democrat	1.000	0.047	8.06e-06	1.83e-04
2024-08-22	3	Anti-Republican	0.153	0.210	4.53e-05	9.43e-04
2024-08-22	3	Others	0.646	0.577	1.06e-04	2.05e-03
2024-08-21	1	Others	0.652	0.577	2.13e-04	3.81e-03

Table 5: Statistically significant changes in party-linked comment proportions before key election events (KEEs).