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

How were deepfakes used before and after the 2024 U.S. presidential election? We present the first quantitative study of this question by analyzing a novel dataset of 231 confirmed deepfakes (images, videos, and audio clips) across social media during the May-December 2024 election period. Our comprehensive statistical analysis of 5 research questions reveals that spikes in deepfake activity preceded key election events (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 led to grave concerns about the use of deepfakes both before and after the 2024 U.S. Presidential election[2]. *Though these past studies offer valuable insights, none of them provides a publicly available dataset of election deepfakes. While they provide insightful qualitative analyses, they do not provide a rigorous statistical analysis. 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. To achieve this, we first carefully curated a dataset of 231 deepfakes (169 images, 38 videos and 24 audios)[3]. 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 support future research, we will 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**: Do KEEs boost engagement (e.g. likes, comments, shares) with deepfakes on social media?
- **RQ5**: Did some KEEs have a greater impact than others at boosting engagement with deepfakes?

Our study sheds light on these questions.[5] Before reading the rest of this paper, we invite you to step back, consider, and then answer these questions for yourself now. Later, you can compare your answers with the findings in the paper.

Related Work

The proliferation of deepfakes (Pei et al. 2024) and their potential impact on societal discourse, particularly in political contexts, has garnered increasing attention in recent years (U.S. Department of Homeland Security 2024; Byman et al.

[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.

[4]This is identical to the ethical use approach proposed by others working on deepfake detection (Li et al. 2020; Dang et al. 2020). For the sake of anonymity, the link to request access to the dataset will be disclosed after the review process.

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

2023; Dalal et al. 2024). 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 A significant body of research has explored the creation, detection, and potential harms of deepfakes across various domains. 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; Chesney and Citron 2019; 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 (Loewenstein 2024; Romero Moreno 2024). However, rapid advances in deepfake creation technology and their increasing availability require continuous monitoring and analysis, particularly in the context of political events like presidential elections. 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.

Temporal Dynamics of Misinformation and Disinformation Understanding the temporal patterns of misinformation and disinformation spread is critical for developing effective mitigation strategies (Hanley and Durumeric 2024). Research in this area has investigated how online rumors and false narratives emerge, propagate, and evolve over time and across platforms often in response to real-world events (Vosoughi, Roy, and Aral 2018; Stella, Ferrara, and De Domenico 2018; Gatta et al. 2023). Studies have shown that significant events can act as catalysts for the rapid dissemination of both accurate and inaccurate information, including manipulated media (Sharma et al. 2019). Our work builds upon this research by specifically examining the temporal relationship between key election events and the publication of deepfakes.

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 (Bakshy et al. 2011; 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 analyzing the social engagement metrics (likes, shares, comments, saves, views) associated with a dataset of deepfakes collected during the 2024 U.S. Presidential Election. By examining the temporal evolution of these metrics and their relationship with election events, we offer 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.

Key Election Events (KEEs)

To identify politically significant events in the 2024 U.S. Presidential election cycle, we consulted the *Ad Fontes Media Bias Chart*[6] and selected three news sources rated as both highly reliable (reliability scores between 0 and 64) and politically neutral (bias scores ranging from -42 [left] to +42 [right])[7]. These sources were the BBC[8] (reliability score: 44.73; bias score: -1.33), Reuters[9] (reliability score: 45; bias score: -1.2), and the Associated Press (AP)[10] (reliability score: 44.82; bias score: -2.38). Key Election Events (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, through 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. This timeline is a reference point for the subsequent analysis of deepfake dissemination during the election period.

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) which is an excellent source of deepfake data, accredited

[6]https://adfontesmedia.com/

[7]Reliability scores above 40 indicate fact-centered analysis; Bias scores between -6 and +6 indicate middle/centric/unbiased reporting.

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

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

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

Table 1: Key Election Events (KEEs) related to the 2024 U.S. Presidential Election. KEEs were extracted from highly-reliable and politically neutral news sources according to the *Ad Fontes Media Bias Chart*.

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			

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[11]. 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 figures (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.

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., Twitter, 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). A datasheet (Gebru et al. 2021) 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 1. The word cloud in Figure 1a was generated by analyzing the titles and textual content of posts featuring deepfakes. Figure 1b illustrates the extent to which major political figures are targeted, while Figure 1c depicts the overall partisan distribution of the dataset. *We underscore that our study is explicitly designed to preserve political neutrality. 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.*

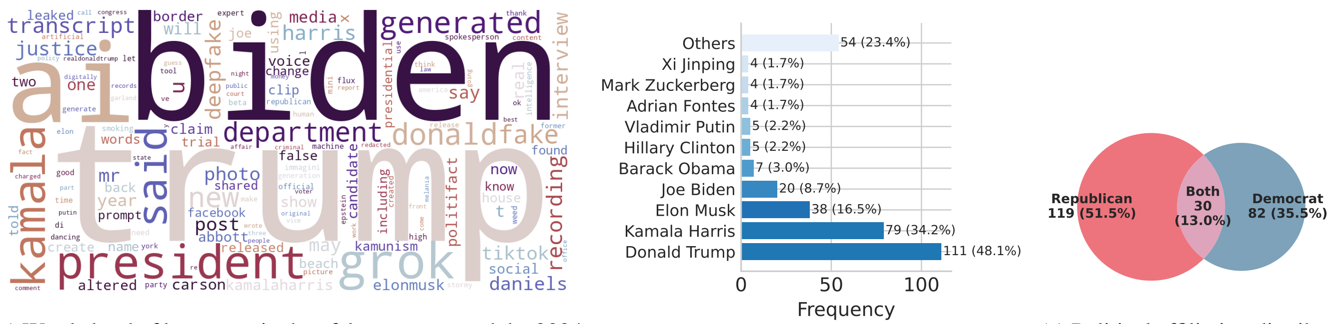
Results

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 1st, 2024 to December 31, 2024. We identified significant spikes in this publication activity and tested for temporal correlations with KEEs.

Figure 2 illustrates the daily number of deepfakes published during the May-Dec 2024 time frame. The dashed vertical lines mark the dates of KEEs, numbered consistently

[11]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).



(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 1: 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).

with Table 2. 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 According to an analysis of z-scores on the daily deepfakes counts, which were then converted to two-tailed p-values using a normal distribution approximation to assess statistical significance, there are six dates with statistically significant spikes ($p < 0.05$) in deepfake publication activity during August 2024. 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 deepfake 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 "key election event 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 FDR correction to control 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 3 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 significant associations. When using a 7-day event window,

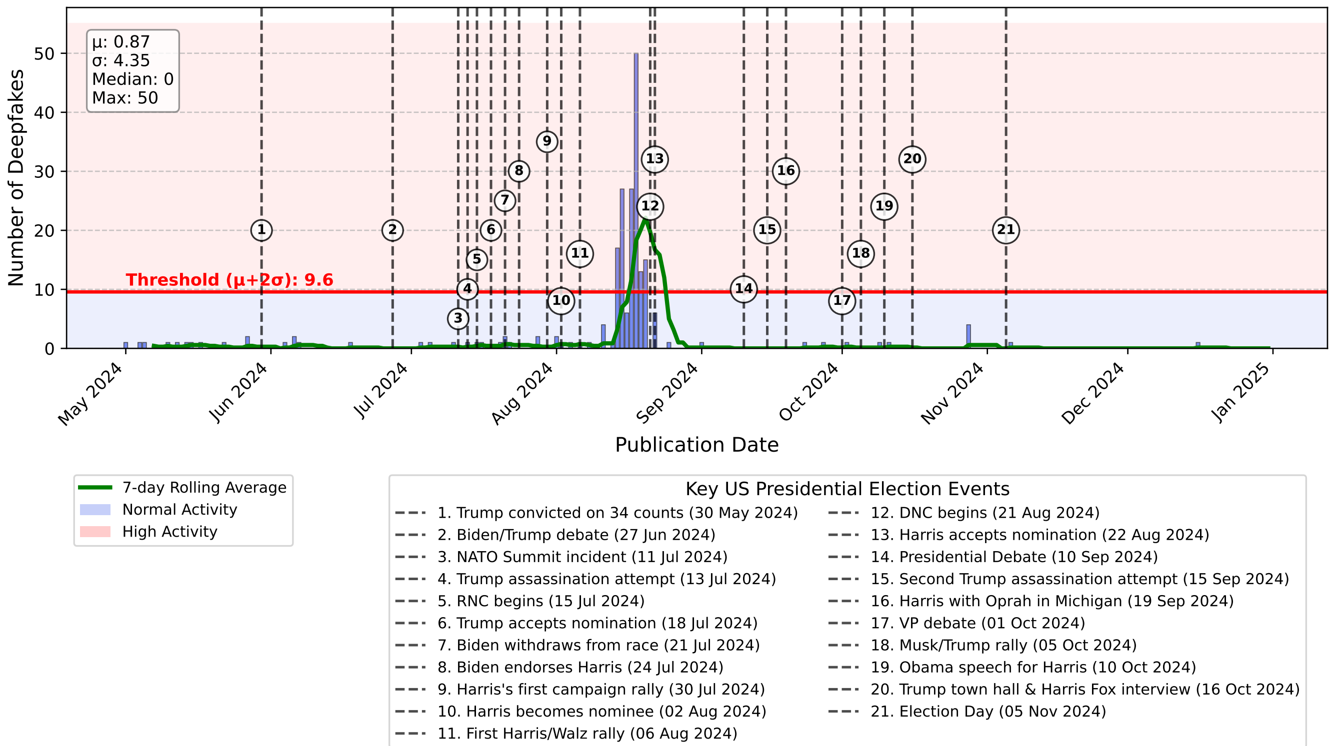


Figure 2: Daily count of deepfakes published around the 2024 U.S. Presidential Election in relation to Key Election Events (KEEs).

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).

> **Takeaway 1:** ☞ *Deepfake activity significantly clustered around key election events (KEEs), especially during spike periods and with wider temporal windows.*

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 4 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.*

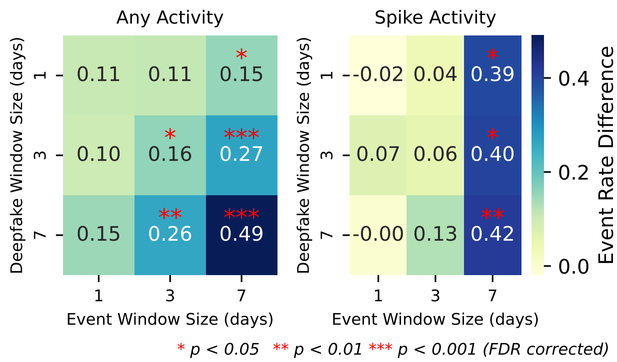


Figure 3: Heatmap showing the difference in Key Election Event (KEE) rates between days near and not near deepfake activity. Color intensity indicates the magnitude of difference, and asterisks (*) mark statistically significant results after FDR correction (* if $p < 0.05$, ** if $p < 0.01$, and *** if $p < 0.001$).

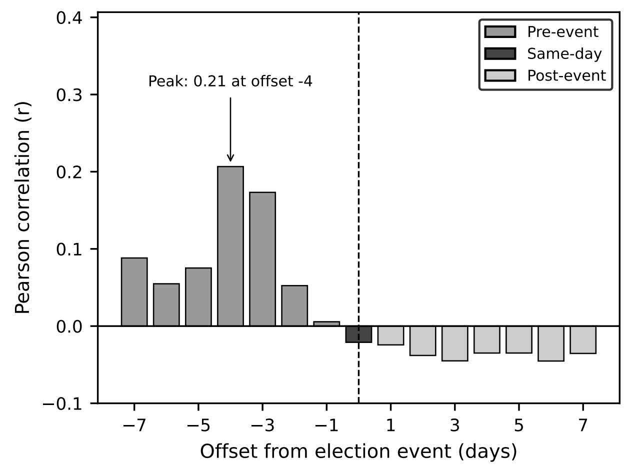


Figure 4: Cross-correlation between daily deepfake publication counts and Key Election Events (KEEs). Correlation coefficients are plotted for temporal offsets from seven days before each event (negative offsets) to seven days after (positive offsets).

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 more 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 5 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 $p > 0.05$).

Takeaway 3: ☞ *Most events related to party conventions and major campaign rallies are linked with significant spikes in deepfake publication activity, whereas debate events showed no significant change in deepfake activity. These KEE-based differences suggest that deepfake creators may strategically target certain types of political events over others.*

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

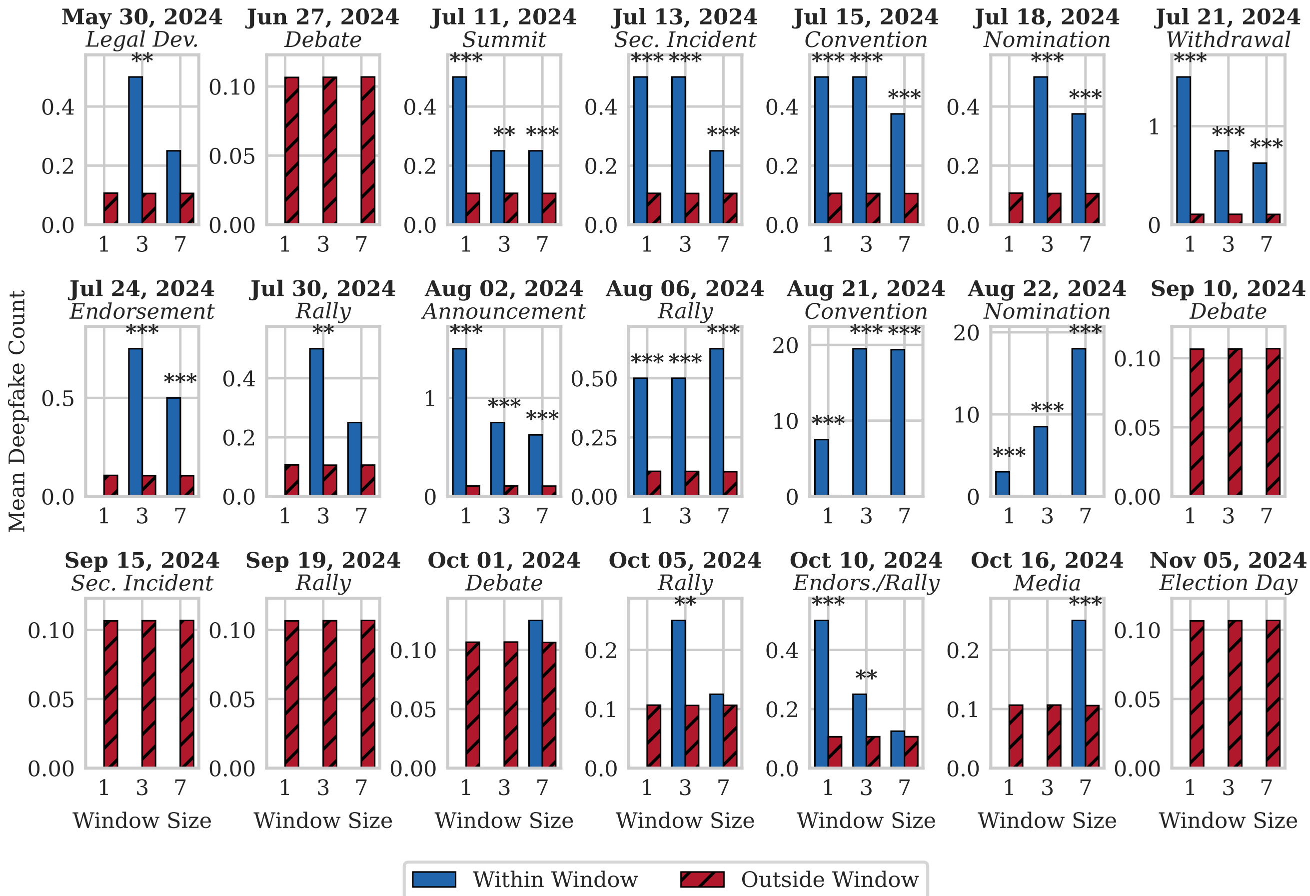


Figure 5: Mean deepfake counts within and outside temporal windows of 1, 3, and 7 days preceding Key Election Events (KEEs). Asterisks denote statistical significance based on Mann-Whitney U tests with Benjamini-Hochberg correction: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Note: y-axis scales vary across histograms.

different engagement metrics evolve over time and whether specific KEEs correlate with higher engagement rates.

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 6 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 4: *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 days in our May-December study period as either falling within these windows (P_0) or outside them (P_1). For each engagement metric, we compared the distributions between these two categories 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.

Table 2 presents the mean values for each engagement metric within and outside the defined temporal windows.

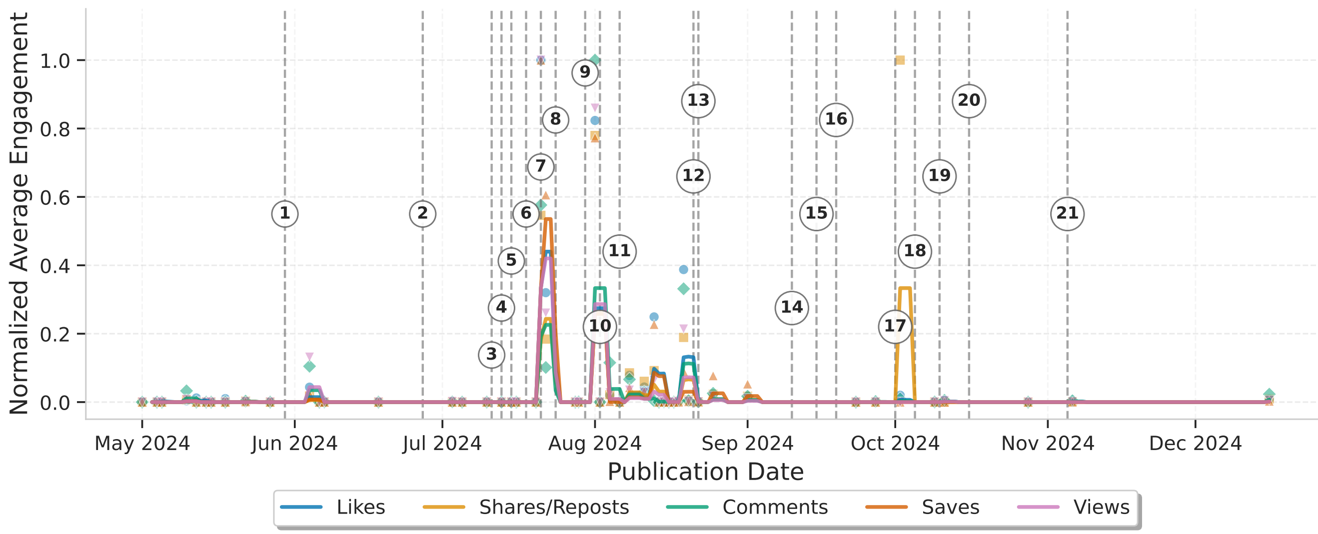


Figure 6: Normalized daily average engagement with deepfakes across five metrics (likes, shares/reposts, comments, saves, and views) from May to December 2024. The vertical dashed lines mark Key Election Events (KEEs), numbered consistently with Table 1. The 3-day rolling averages (solid lines) highlight general trends in engagement patterns.

Table 2: Average engagement (likes, shares/reposts, comments, saves, views) for deepfakes published within temporal windows preceding KEEs (population P_0) versus days outside these windows (population P_1). Window sizes of 1, 3, and 7 days (ω) were analyzed for each metric.

	$\omega = 1$		$\omega = 3$		$\omega = 7$	
Engagement 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

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 ($p = 5.41 \times 10^{-6}$, 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 5: *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 6, 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.

Takeaway 5: ☞ *Engagement with deepfakes did not exhibit statistically significant changes before individual key election events (KEEs), except for a notable decrease preceding one event.*

Discussion

Our analysis reveals clear associations between deepfake publication activity and key election events (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. The Democratic National Convention and related events (August 21-22) triggered huge spikes in deepfake activity, 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.

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 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 rather than uniform distribution throughout the election cycle. The anticipatory nature of deepfake production indicates that proactive monitoring and rapid response systems *before* KEEs may be more effective than post-hoc fact-checking.

Limitations We acknowledge several limitations in our research. First, despite our multi-source collection approach (PDID, fact-checking organizations, and Google Alerts) and manual screening process, platform or topic-specific biases may remain, and the dataset may not capture all deepfakes circulated during the 2024 U.S. Presidential Election. Deepfakes in closed networks, private messaging applications, or those not flagged by fact-checkers likely remain undocumented. Second, while we identify temporal correlations between KEEs and deepfake activity, our methodology cannot definitively establish causality. These correlations may be interpreted through competing hypotheses—for example, that deepfake spikes reflect increased public attention around election-related events or are part of strategic communication efforts by malign actors. In addition, alternative mechanisms could also account for the observed outcomes, such as algorithmic amplification by social media platforms, content recommendation patterns, or coincidental spikes in user engagement unrelated to KEEs. Third, while our analysis examines various temporal windows (1, 3, and 7 days), these discrete intervals may not fully capture the dynamic nature of online content dissemination. Certain deepfakes might have delayed effects or longer engagement lifecycles that extend beyond our analyzed windows.

Ethical Statement Throughout our research process, we have adhered to stringent ethical standards. We have ensured that no personally identifiable information was utilized for non-public individuals, and all analyses were conducted on aggregated data. Our research has been reviewed by an Institutional Review Board (IRB)[12] which determined that the activity is not research involving human subjects. Additionally, we have thoroughly considered the societal impacts of our research. We acknowledge that documenting the relationship between key election events (KEEs) and deepfake publication activity, as well as engagement patterns, may potentially lead to misuse of our research and/or to maximize the impact of synthetic media campaigns. To mitigate this risk, we have deliberately presented our findings at a level of granularity that provides useful insights without offering a tactical playbook for attackers, and we restrict access to our dataset through an ethical usage policy.

[12]For the sake of anonymity, we will provide information on the University and the IRB Approval Number after the review process.

Conclusion & Future Work

This study presents the first statistical analysis of deepfake activity in the 2024 U.S. presidential election, revealing patterns in the timing, dissemination, and public engagement with synthetic media during a high-stakes political process. While the dataset provides unprecedented insights into the evolving threat of AI-generated disinformation, it also highlights the need for platform policies and detection systems that account for event-driven surges in synthetic media. This study underscores the urgency of cross-sector collaboration among researchers, platforms, and policymakers to preserve electoral integrity in the age of generative AI.

Future work should expand to global elections to examine how geopolitical, cultural, and regulatory differences influence the volume, timing, and impact of election-related deepfakes.

Our research provides some simple guidelines for Regulatory Entities[13] (REs) that seek to minimize harmful effects of deepfakes during elections. First, dates of many Key Election Events are known well in advance — in such cases, REs should ramp up their deepfake detections processes (including both automated and human moderators) around the KEE dates. Second, we recommend that REs should always have "surge teams" available on short notice when an unexpected KEE occurs (such as the attempted assassination of President Trump). Such surge teams should instantly jump into action when an unexpected KEE occurs and ensure that they can heighten efforts to proactively detect deepfakes when they are posted, rather than well after the fact.

[13]Regulatory Entities (REs) might include government entities charged with regulating mis/disinformation, as well as social platforms' organizations charged with the same responsibilities. In fact, they could also include third party academic entities and/or NGOs that are monitoring the integrity of elections — though they may not have regulatory authority, they can alert such REs.

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.

Bakshy, E.; Hofman, J. M.; Mason, W. A.; and Watts, D. J. 2011. Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining*, 65–74.

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.

Chesney, R.; and Citron, D. 2019. Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. *Foreign Aff.*, 98: 147.

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.

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).

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.

Hanley, H. W.; and Durumeric, Z. 2024. Machine-made media: Monitoring the mobilization of machine-generated articles on misinformation and mainstream news websites. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18, 542–556.

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.

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

Stella, M.; Ferrara, E.; and De Domenico, M. 2018. Bots increase exposure to negative and inflammatory content in online social systems. *Proceedings of the National Academy of Sciences*, 115(49): 12435–12440.

U.S. Department of Homeland Security. 2024. Increasing Threats of Deepfake Identities. Accessed: 2024-12-13.

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.

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.

Ethics 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.

USPED2024 Datasheet

Motivation

For what purpose was the dataset created? *Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.* The 2024 U.S. Presidential Election Deepfakes Dataset (USPED) was created to address a research gap in the study of deepfakes in electoral contexts. While previous studies have documented instances of synthetic media in various elections worldwide, none offered both a comprehensive, publicly available dataset and rigorous quantitative analysis of deepfake dynamics. This dataset was specifically designed to investigate the relationship between Key Election Events (KEEs) and deepfake activity through statistical analysis of publication patterns and engagement metrics. By documenting 231 carefully curated deepfakes (169 images, 38 videos, and 24 audios) from the 2024 U.S. Presidential Election cycle, this resource enables researchers to empirically examine how synthetic media interacts with significant political moments. The dataset will be made available to academics who agree to an ethical usage policy, supporting future research on the potential impact of deepfakes on democratic processes.

Who created the dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)? For the sake of anonymity, these details will be disclosed after the review process.

Who funded the creation of the dataset? *If there is an associated grant, please provide the name of the grantor and the grant name and number.* For the sake of anonymity, these details will be disclosed after the review process.

Any other comments? None.

Composition

What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)? *Are there multiple types of instances (e.g., movies, users, and ratings; people and interactions be tween them; nodes and edges)? Please provide a description.* Each instance in the dataset represents a specific deepfake media item that was carefully curated and annotated with metadata such as the source URL, publication date, subject name (political figure depicted), content category, language, engagement metrics (likes, comments, reposts/shares, saves, views), platform source, and other relevant attributes. The deepfakes in the dataset specifically depict or reference political figures associated with the 2024 U.S. Presidential Election, including candidates, elected officials, news anchors, and politically associated public figures. The dataset does not contain the actual deepfake content.

How many instances are there in total (of each type, if appropriate)? The dataset contains 231 deepfakes in total: 169 images (visual deepfakes), 38 videos (visual deepfakes with motion) and 24 audios. Videos with audio and visual components are counted independently as being both a video and an audio sample.

Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? *If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (e.g., geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (e.g., to cover a more diverse range of instances, because instances were withheld or unavailable).* The dataset is a sample, not an exhaustive collection, of deepfakes related to the 2024 U.S. Presidential Election. The larger set would encompass all deepfakes created and shared across all platforms (public and private) in relation to the 2024 U.S. Presidential Election. While we employed a multi-faceted collection strategy—drawing from the Political Deepfakes Incidents Database (PDID), established fact-checking organizations (AFP Fact Check, AI Incident Database, FactCheck.org, OpenSecrets, PolitiFact, and Snopes), and Google Alerts—this approach inherently captures only a subset of the total deepfakes circulated during this period. Deepfakes in closed networks, private messaging applications, or those not flagged by fact-checkers likely remain undocumented. Additionally, the coverage focuses on the U.S. Presidential Election context, which means that the dataset is not representative of global deepfake activity. The temporal coverage spans from May 1, 2024, to December 31, 2024, capturing the pre-election campaign period through post-election discourse.

What data does each instance consist of? *"Raw" data (e.g., unprocessed text or images) or features? In either case, please provide a description.* Each instance in the dataset consists extensive metadata features associated with deepfakes. Specifically, we recorded the following attributes: Deepfake source, URL, Publish date, Source text/title, Subject name (the political figure depicted), Category, Language, Media type (video, audio, or image), Online accessibility status, Engagement metrics (i.e. Number of likes, Shares/reposts, Comments, Saves, Views), Misinformation warning flags (e.g., fact-check labels applied by platforms), Original platform (e.g., Twitter, YouTube), Original collection source (PDID, fact-checkers, or Google Alerts), Inferred political affiliation of the subject (Democrat, Republican, or both).

Is there a label or target associated with each instance? *If so, please provide a description.* No.

Is any information missing from individual instances? *If so, please provide a description, explaining why this information is missing (e.g., because it was unavailable). This does not include intentionally removed information, but might include, e.g., redacted text.* Yes. Information may be missing in cases where attributes were unavailable or not applicable. For example, information about re-posts might be missing when deepfakes have been published on social platforms that do not provide the re-post functionality.

Are relationships between individual instances made explicit (e.g., users' movie ratings, social network links)?

If so, please describe how these relationships are made explicit. None explicitly, though the data samples include the URL to the posted deepfake, so some information (such as threads, replies, or posts by the same author) could be extracted if needed.

Are there recommended data splits (e.g., training, development/validation, testing)? *If so, please provide a description of these splits, explaining the rationale behind them.* Not applicable.

Are there any errors, sources of noise, or redundancies in the dataset? *If so, please provide a description.* See preprocessing below.

Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)? *If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (i.e., including the external resources as they existed at the time the dataset was created); c) are there any restrictions (e.g., licenses, fees) associated with any of the external resources that might apply to a dataset consumer? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.* The dataset is self-contained for its primary research purpose—analyzing deepfake dissemination patterns and engagement metrics around election events. The core dataset includes all necessary metadata (e.g., timestamps, media types, annotations, and statistical features) to support reproducibility without dependency on external resources. However, we note that supplemental links to original deepfake content (e.g., social media posts) are provided for researchers interested in examining the media itself. These links are secondary to the study's focus and may become unavailable over time due to platform removals.

Does the dataset contain data that might be considered confidential (e.g., data that is protected by legal privilege or by doctor patient confidentiality, data that includes the content of individuals' non-public communications)? *If so, please provide a description.* No.

Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? *If so, please describe why.* The dataset does not contain data that is directly offensive, insulting, or threatening. However, it is important to note that the dataset includes URLs linking to actual deepfake content involving public figures from the U.S. 2024 Presidential Election cycles. While these links are provided for transparency and research purposes, the deepfakes themselves may be considered offensive or distressing, particularly by supporters of the individuals depicted.

Does the dataset identify any subpopulations (e.g., by age, gender)? *If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.* No.

Is it possible to identify individuals (i.e., one or more natural persons), either directly or indirectly (i.e., in combination with other data) from the dataset? *If so, please describe how.* Yes, the names of the public figures associated with each deepfake in our dataset can be identified.

Does the dataset contain data that might be considered sensitive in any way (e.g., data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? *If so, please provide a description.* No.

Any other comments? No.

Collection Process

How was the data associated with each instance acquired? *Was the data directly observable (e.g., raw text, movie ratings), reported by subjects (e.g., survey responses), or indirectly inferred/derived from other data (e.g., part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.* The data for each deepfake instance was acquired through direct observation of publicly available content from curated sources like the Political Deepfakes Incidents Database (PDID), fact-checking organizations (e.g., PolitiFact, Snopes), and Google Alerts. Two trained analysts manually verified each sample, excluding irrelevant or duplicate content, and recorded metadata (e.g., URLs, engagement metrics, subject names) with direct observation. Political affiliations were inferred, and misinformation warnings were noted if flagged by platforms. While engagement metrics were taken as reported, the analysts cross-verified attributes to ensure consistency.

What mechanisms or procedures were used to collect the data (e.g., hardware apparatuses or sensors, manual human curation, software programs, software APIs)? *How were these mechanisms or procedures validated?* The data was collected through manual human curation, including trained analysts reviewing deepfakes from fact-checking platforms (PolitiFact, Snopes), the Political Deepfakes Incidents Database (PDID), and Google Alerts monitoring for election-related synthetic media mentions. Validation occurred through a dual-analyst verification system to ensure labeling consistency, cross-referencing with accredited fact-checking organizations to confirm authenticity, and ethical screening to exclude unverifiable or private content.

If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)? The dataset represents a non-probabilistic, purposive sample drawn from a larger universe of potential deepfake content, with collection guided by temporal coverage (May–December 2024 to align with the election cycle) and political relevance (only content depicting U.S. presidential candidates, officials, or closely associated figures).

Who was involved in the data collection process (e.g., students, crowdworkers, contractors) and how were they compensated (e.g., how much were crowdworkers paid)? Authors of the original paper which released this dataset have been involved in the data collection process.

Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (e.g., recent crawl of old news articles)? *If not, please describe the timeframe in which the data associated with the instances was created.* Data collection occurred from May 1st, 2024 to May 1st, 2025. This timeframe overlaps with the creation timeframe of the data.

Were any ethical review processes conducted (e.g., by an institutional review board)? *If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.* Yes. Our data collection process has been reviewed by an Institutional Review Board[14], which determined that the proposed activity was not research involving human subjects.

Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? *If so, please provide a description. If not, you may skip the remaining questions in this section.* Yes, the dataset underwent systematic preprocessing and labeling to ensure consistency and quality. The raw data was cleaned by removing duplicates through manual verification by two analysts, who cross-checked URLs, content hashes, and metadata. Each deepfake instance was labeled with standardized attributes (e.g., media type, subject name, political affiliation) based on analyst review and fact-checker annotations. Additionally, content filtering was applied to exclude cases unrelated to the 2024 U.S. Presidential Election cycle.

Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)? *If so, please provide a link or other access point to the "raw" data.* No, the raw data was not preserved separately to prevent potential misuse and dissemination of unverified or unredacted content. To maintain ethical standards and minimize risks of harmful repurposing, only the curated, preprocessed dataset—with strict access controls—is made available to qualified researchers under an ethical use agreement.

Is the software that was used to preprocess/clean/label the data available? *If so, please provide a link or other access point.* Not applicable (manual human curation).

Any other comments? None.

Uses

[14]For the sake of anonymity, we will provide information on the University and the IRB Approval Number after the review process.

Has the dataset been used for any tasks already? *If so, please provide a description.* Yes, the dataset has been used exclusively for the analyses presented in its related study, specifically to: (1) quantify temporal relationships between deepfake dissemination and key election events, (2) measure engagement patterns across social media platforms, and (3) identify event-specific dynamics in synthetic media proliferation.

Is there a repository that links to any or all papers or systems that use the dataset? *If so, please provide a link or other access point.* No.

What (other) tasks could the dataset be used for? This dataset could enable: (1) detection model benchmarking (testing AI classifiers on real-world election deepfakes), (2) narrative analysis (studying manipulative themes tied to specific candidates/events), (3) platform policy audits (evaluating how social media enforce synthetic media rules during elections), and (4) longitudinal studies of deepfake evolution across election cycles. However, access restrictions will apply to prevent misuse—all proposed uses must align with ethical guidelines and demonstrate non-harmful intent. Researchers may apply for dataset access solely through controlled academic channels.

Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? *For example, is there anything that a dataset consumer might need to know to avoid uses that could result in unfair treatment of individuals or groups (e.g., stereotyping, quality of service issues) or other risks or harms (e.g., legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to mitigate these risks or harms?* The dataset's composition and collection methods introduce some limitations for future use, including potential selection bias from prioritizing fact-checker-identified and viral content (which may overrepresent deepfakes targeting prominent candidates), possible distortion from platform engagement metrics that reflect algorithmic amplification rather than organic reach, and exclusion of unverified fringe content that could reveal broader dissemination patterns. To mitigate these issues, researchers using this dataset should explicitly acknowledge these biases in any analyses, supplement findings with platform transparency data when available, avoid making definitive partisan attributions without additional evidence, and adhere to strict ethical guidelines that will be enforced through controlled access protocols requiring formal acknowledgment of these limitations in all resulting publications.

Are there tasks for which the dataset should not be used? *If so, please provide a description.* This dataset should not be used for: (1) developing real-time deepfake generation tools, (2) partisan attribution without independent verification (due to potential selection biases), or (3) any application that violates the ethical use policy (e.g., targeted harassment of depicted individuals). The temporal and engagement patterns are specific to the 2024 election context and should not be extrapolated to other elections without robustness checks.

Dataset access will be denied for projects lacking safeguards against these misuse scenarios.

Any other comments? None.