

# Tracking the 2024 US Presidential Election Chatter on TikTok: A Public Multimodal Dataset

Gabriela Pinto, Charles Bickham, Tanishq Salkar, Joyston Menezes, Luca Luceri, Emilio Ferrara

gpinto,cbickham,salkar,jmenezes,lluceri,emiliofe@usc.edu

University of Southern California

Los Angeles, California, USA

## Abstract

This paper presents the TikTok 2024 U.S. Presidential Election Dataset, a large-scale, resource designed to advance research into political communication and social media dynamics. The dataset comprises 3.14 million videos published on TikTok between November 1, 2023, and October 16, 2024, encompassing video ids and transcripts. Data collection was conducted using the TikTok Research API with a comprehensive set of election-related keywords and hashtags, supplemented by third-party tools to address API limitations and expand content coverage, enabling analysis of hashtag co-occurrence networks that reveal politically aligned hashtags based on ideological affiliations, the evolution of top hashtags over time, and summary statistics that highlight the dataset’s scale and richness. This dataset offers insights into TikTok’s role in shaping electoral discourse by providing a multimodal view of election-related content. It enables researchers to explore critical topics such as coordinated messaging, misinformation spread, audience engagement, and linguistic trends. The TikTok 2024 U.S. Presidential Election Dataset is publicly available and aims to contribute to the broader understanding of social media’s impact on democracy and public opinion.

## Keywords

U.S. Presidential Election, TikTok, social media

### ACM Reference Format:

Gabriela Pinto, Charles Bickham, Tanishq Salkar, Joyston Menezes, Luca Luceri, Emilio Ferrara. 2018. Tracking the 2024 US Presidential Election Chatter on TikTok: A Public Multimodal Dataset. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym ’XX)*. ACM, New York, NY, USA, 4 pages. <https://doi.org/XXXXXXX.XXXXXXX>

## 1 Introduction

Social media has profoundly reshaped electoral politics, emerging as a critical medium for disseminating election-related information and influencing public opinion. This shift has garnered significant attention from researchers [1, 5–8]. Although platforms like Twitter have long been a cornerstone in the study of global geopolitical

events [2–4], TikTok offers a distinctive perspective on political discourse. Its short-form video format, combined with interactive features like comments, hashtags, and trends, enables rapid content dissemination and highly visual engagement, making it uniquely suited for studying political communication. With an estimated global user base exceeding one billion,<sup>1</sup> TikTok has quickly become a platform where political campaigns engage and inform audiences, particularly adolescents [9].

The 2024 U.S. Presidential Election highlighted TikTok’s growing influence in electoral politics, as major candidates actively engaged with the platform to connect with voters. For instance, Donald Trump joined TikTok in June 2024, rapidly gaining over 6 million followers,<sup>2</sup> while Joe Biden actively posted over 200 videos starting in February 2024, amassing more than 373,000 followers.<sup>3</sup> These examples illustrate TikTok’s potential as both a campaign tool and a rich data source for understanding electoral sentiment, political narratives, and voter outreach strategies. Its unique format and audience demographic make TikTok an important platform for studying the evolving nature of political communication in the digital age.

This paper introduces the TikTok 2024 U.S. Presidential Election Dataset, a collection of over 3 million videos that includes meta-data and textual transcripts. The dataset captures the dynamics of TikTok’s election-related discourse, enabling researchers to explore key questions such as identifying coordinated messaging among political actors, detecting the spread of misinformation, examining audience engagement patterns, and analyzing discourse across linguistic and demographic groups.

By leveraging this dataset, researchers can gain a deeper understanding of how political communication evolves in digital spaces, particularly on TikTok’s unique platform. It highlights TikTok’s growing influence in shaping public opinion, especially among younger voters—a pivotal demographic in modern election. Ultimately, this dataset seeks to contribute to broader discussions about the role of digital platforms in electoral processes and the implications for digital democracy in the age of social media.

## 2 Method of Data Collection

### TikTok API and Data Collection

The TikTok Research API<sup>4</sup> provides access to both account-related and content-related data on TikTok. Account data includes user profiles, follower lists, liked videos, and reposts, while content data encompasses video descriptions, hashtags, comments, subtitles, and

<sup>1</sup><https://whatsthebigdata.com/tiktok-statistics/>

<sup>2</sup><https://cnn.com/2024/06/02/politics/donald-trump-joins-tiktok/index.html>

<sup>3</sup><https://cnn.com/2024/06/12/tech/tiktok-pew-research-politics-x/index.html>

<sup>4</sup><https://developers.tiktok.com/doc/research-api-specs-query-videos/>

engagement metrics such as likes, shares, and comments. Accessing the API requires a bearer token, generated using client and secret keys, which must accompany API requests. After exploring third-party scrapers, the Official Research API proved to be the most reliable method for metadata collection.

**Video Metadata Collection.** We used the TikTok Research API to collect metadata for videos, leveraging parameterized queries to target content by regions, topics, languages, or publication dates. Key parameters included the date range, defining the start and end dates for data collection, and fields, which specified video ID, description, creation time, region, engagement metrics (likes, views, comments, shares), music ID, hashtags, and username. Queries were restricted to 100 records per call, and advanced filters, using logical operators (e.g., AND/OR), allowed precise targeting of content based on specific hashtags and regions. For instance, videos tagged with #election2024 in the U.S. were queried using region\_code="US" and hashtag="#election2024".

The metadata collection process was structured around phases defined by significant political events, with the project divided into several phases spanning specified date ranges. Phase 1 spans from November 1, 2023, to January 15, 2024. Phase 2 continues from January 16, 2024, to February 7, 2024, followed by Phase 3, which occurs between February 8, 2024, and February 27, 2024. Phase 4 is scheduled from February 28, 2024, to March 5, 2024. Subsequently, Phase 5 runs from March 6, 2024, to July 14, 2024, and Phase 6 takes place between July 15, 2024, and July 20, 2024. Phase 7 follows shortly after, from July 20, 2024, to August 5, 2024. Phase 8 is set for the period of August 6, 2024, to September 9, 2024, and Phase 9 will occur between September 10, 2024, and September 15, 2024. Phase 10 is scheduled from September 15, 2024, to November 10, 2024, and finally, Phase 11 begins on November 11, 2024, and continues to the present.

## Video Downloads and Transcript Generation

To complement the collected metadata, we used PykTok<sup>5</sup>, a Python library designed to interact with TikTok's platform, to download the corresponding video content. After downloading, we utilized OpenAI Whisper<sup>6</sup> to generate transcripts for each video. To ensure accuracy, we randomly selected a subset of 100 videos and verified the correctness of their generated transcripts.

**Note:** The data collection process complies fully with TikTok's Terms of Service, ensuring adherence to ethical standards.

## 3 Exploratory Analysis

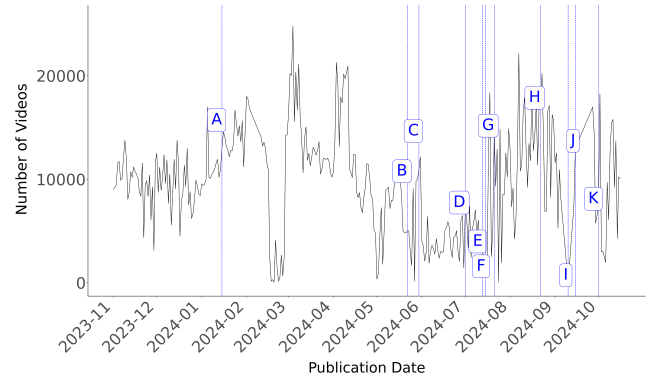
Using the collected data, we performed an exploratory analysis to identify key patterns and trends in TikTok's election-related discourse. This analysis offers an initial glimpse into how users engaged with and discussed major political events and candidates during the 2024 U.S. Presidential Election.

### Summary of Data

Table 1 summarizes the statistics for the current dataset. TikTok-generated transcripts account for only 15.4% of the total videos, reflecting a limitation of the TikTok Research API. To enrich the

Number of videos	3,144,836
Number of transcripts	485,343
Number of comments	161,892,802
Number of views	18,010,334,100
Number of likes	2,042,078,329
Number of shares	119,934,718
Number of unique users	693,826

**Table 1: Summary statistics of the dataset.**



**Figure 1: Timeline of events and volume of TikTok posts.**

dataset with more linguistic data, we incorporate Whisper-generated transcripts to our dataset and related GitHub repository.

### Number of videos over time

This dataset includes video metadata collected for posts published between November 1, 2023, and October 16, 2024. Figure 1 shows the number of videos collected by publication date, along with key political events during the election cycle and the corresponding number of posts on those dates. The data collection process will conclude after the Inauguration Day in January 2025.

### Top Hashtags

Table 2 presents the 20 most frequent hashtags for each video. Each video collected through the TikTok Research API provides

<sup>5</sup><https://github.com/dfreelon/pyktok>

<sup>6</sup><https://github.com/openai/whisper>

Tag	Count
trump	774,506
trump2024	465,388
maga	349,150
biden	338,506
republican	263,289
donaldrump	261,756
usa	234,190
politics	204,580
democrat	181,544
america	175,687
kamalaharris	157,783
joebiden	150,204
news	146,906
voteblue	144,656
election	141,000
democrats	135,433
conservative	134,280
vote	126,958
kamala	115,919
election2024	109,770

Table 2: Top 20 Most Frequently Occurred Hashtags.

a list of hashtags labeled “hashtag\_names” Overall, the hashtags reflect that the videos labeled by the creator were related to Donald Trump, President Joe Biden, and the right-leaning ideology through keywords such as “maga” (“Make America Great Again”) and “republican”

Language Detection

We used LangDetect<sup>7</sup> to identify the languages of the transcripts generated by the TikTok Research API. In our current dataset, 460,735 transcripts have been classified as English-based content, making it the dominant language. The second most common language is Spanish, with 15,621 transcripts. In future work, we plan to analyze Spanish-based content further to provide a more comprehensive view of the discourse surrounding the election cycle.

Hashtag Co-Occurrence Graph

Figure 2 shows a hashtag co-occurrence graph of TikTok content related to the 2024 U.S. Election. We selected the top 35 left-leaning and right-leaning hashtags, including additional hashtags that co-occurred with them. Hashtag pairs appearing together more than 500 times were retained, and the network was visualized in Gephi<sup>8</sup> using the modularity class algorithm, revealing a political divide.

The Right-Leaning cluster (Red) includes interconnected hashtags like #trumpsupporters, #trumptrain, #trump, #donaldrump, and #maga, reflecting strong support for Trump and Republican themes. The Left-Leaning cluster (Blue) features hashtags such as #kamalaharris, #voteblue, #voteblue2024, #bluewave, and #biden, emphasizing Democratic themes and voter mobilization.

Hashtags Over Time

Figure 3 reveals significant trends across political and neutral categories. Republicans accounted for the majority of hashtags, totaling 1,882,321, with a notable peak of 18,045 hashtags on July 18, 2024.

<sup>7</sup><https://pypi.org/project/langdetect/>  
<sup>8</sup><https://gephi.org/>

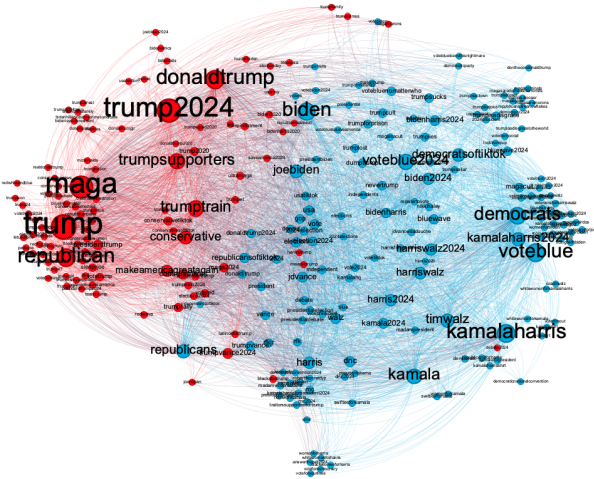


Figure 2: Hashtag Co-Occurrence Graph

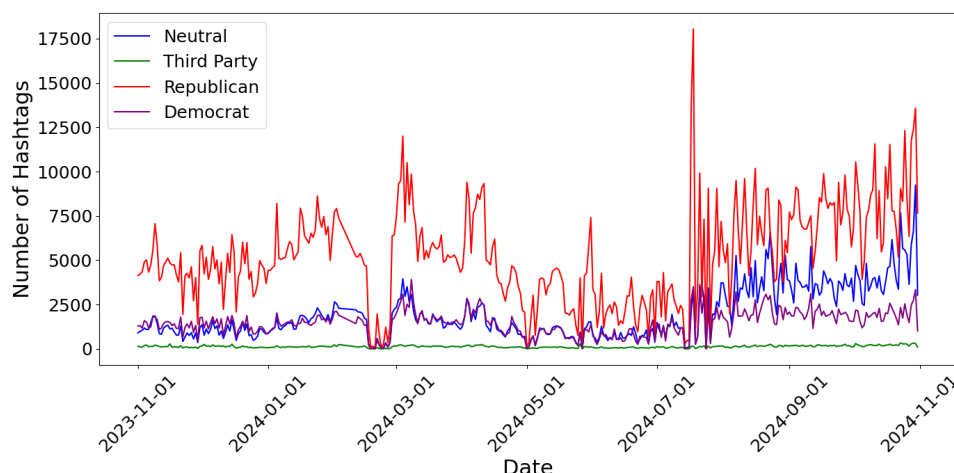
Neutral hashtags were the second most common, reaching a peak of 9,243 on October 30, 2024, and accumulating 696,917 overall. Democrats followed with a total of 524,664 hashtags, peaking at 3,918 on March 8, 2024. Third-party hashtags were minimal at 44,980 total, peaking at 336 on October 23, 2024. The “Other” category saw an extraordinary spike on March 4, 2024, with 243,128 hashtags. This data underscores the dynamic engagement trends across political affiliations and neutral discussions, with Republicans dominating both overall volume and peak activity periods.

Topic Modeling

To analyze recurring themes related to the 2024 US Presidential election, we conducted topic modeling and summarization of TikTok transcripts from April to November 2024. Using BERTopic for topic extraction and the LLaMA 3.1 8B-Instruct model for summarization, we examined weekly discussions, filling in missing data with Whisper. Immigration emerged as a key theme, with bilingual debates on unresolved issues at the US-Mexico border and dissatisfaction with government responses. Social issues such as gun violence, systemic corruption, reproductive rights, and economic struggles were also prominent, reflecting widespread user concerns. Religion featured in discussions about Catholic and Muslim values, religious texts, and their implications for politics.

Conspiracy theories were prevalent, including speculative discussions about security breaches, assassination attempts, and media bias, with concerns over misinformation frequently highlighted. International conflicts, such as the Israel-Palestine conflict and the Ukraine war, also drew attention, particularly regarding the US role in these crises. Political figures dominated many conversations, with Donald Trump’s legal challenges, policies, and reelection prospects sparking mixed sentiments, while Joe Biden faced criticism over





**Figure 3: Hashtag usage trends by category from November 2023 to October 2024, showing peak activity for Republicans, Democrats, Neutral, and Third-Party hashtags over time.**

immigration laws, inflation, and rumors of withdrawal from the race. Kamala Harris garnered discussions about her fundraising efforts, policies, and potential to become the first female US president. This analysis provides a comprehensive view of the diverse political and social narratives shaping TikTok conversations during the 2024 election cycle.

## 4 Conclusions

The TikTok 2024 U.S. Presidential Election Dataset offers a large-scale resource for advancing research into political communication and social media dynamics. Spanning 3.14 million videos collected using the TikTok Research API and third-party tools, it provides a comprehensive view of election-related content from November 1, 2023, to October 16, 2024. This dataset enables the analysis of critical topics such as misinformation spread, audience engagement, and coordinated messaging, while also capturing hashtag co-occurrence networks and trends over time. By making this dataset publicly available, we aim to empower researchers to explore TikTok's role in shaping electoral discourse and its broader impact on democracy and public opinion.

## Limitations

The TikTok Research API imposes constraints such as rate limits, token expiration, and daily request caps, which made large-scale data collection time-intensive. To mitigate these challenges, we prioritized video metadata collection and expanded our team to distribute the workload, ensuring more efficient data retrieval.

## Future Work

Data collection will conclude a week after Inauguration Day in January 2025. To analyze the online discourse surrounding the election and its aftermath, we will prioritize collecting comments and replies from recent months (starting in May) and work backward as the process becomes more efficient. Our public repository will be continuously updated with video IDs, detected transcript languages, comment summaries, and AI-generated video descriptions.

This evolving dataset will enable meaningful analysis of public sentiment and discussion trends related to the election.

## Data Availability

The dataset, available via GitHub, complies with TikTok's Terms and Conditions, which restrict the direct distribution of collected videos. To adhere to these guidelines, we provide Video IDs—unique identifiers for each post—rather than the videos themselves. These IDs enable researchers to use TikTok's API to retrieve complete video objects, including associated multimedia and metadata. The repository, that is available at <https://github.com/gabbypinto/US2024PresElectionTikToks>, will be continuously updated to ensure the dataset remains current and accessible.

## References

- [1] Anton Abilov, Yiqing Hua, Hana Matatov, Ofra Amir, and Mor Naaman. 2021. Voterfraud2020: a multi-modal dataset of election fraud claims on twitter. In *Proc. Int. AAAI Conference on Web & Social Media*. 901–912.
- [2] Emily Chen, Ashok Deb, and Emilio Ferrara. 2022. #Election2020: the first public Twitter dataset on the 2020 US Presidential election. *Journal of Computational Social Science* (2022), 1–18.
- [3] Emily Chen and Emilio Ferrara. 2023. Tweets in Time of Conflict: A Public Dataset Tracking the Twitter Discourse on the War Between Ukraine and Russia. In *Proc. 17th Int. AAAI Conference on Web & Social Media*. 1006–1013.
- [4] Clayton A Davis, Giovanni Luca Ciampaglia, Luca Maria Aiello, Keychul Chung, Michael D Conover, Emilio Ferrara, Alessandro Flammini, Geoffrey C Fox, Xiaoming Gao, Bruno Gonçalves, et al. 2016. OSoMe: the IUNI observatory on social media. *PeerJ Computer Science* 2 (2016), e87.
- [5] Ashok Deb, Luca Luceri, Adam Badawy, and Emilio Ferrara. 2019. Perils and Challenges of Social Media and Election Manipulation Analysis: The 2018 US Midterms. In *Proc. 2019 World Wide Web Conference*. 237–247.
- [6] Andreas Jungherr. 2016. Twitter use in election campaigns: A systematic literature review. *Journal of information technology & politics* 13, 1 (2016), 72–91.
- [7] Nane Kratzke. 2017. The #btw17 Twitter dataset—recorded tweets of the federal election campaigns of 2017 for the 19th German Bundestag. *Data* 2, 4 (2017), 34.
- [8] Luca Luceri, Ashok Deb, Silvia Giordano, and Emilio Ferrara. 2019. Evolution of bot and human behavior during elections. *First Monday* 24, 9 (2019).
- [9] Christian Montag, Haibo Yang, and Jon D Elhai. 2021. On the psychology of TikTok use: A first glimpse from empirical findings. *Frontiers in public health* 9 (2021), 641673.