

Tracking Political Sentiments: Analyzing Election-Driven Discussions on Reddit and 4chan

Asmita Singh
Binghamton University
Binghamton, NY, USA
asingh14@binghamton.edu

Chinmay Purandare
Binghamton University
Binghamton, NY, USA
cpurandare@binghamton.edu

Sanika Shinde
Binghamton University
Binghamton, NY, USA
sshinde2@binghamton.edu

Abstract

This report investigates the online discourse surrounding the 2024 U.S. presidential election on platforms like Reddit and 4chan, focusing on utilizing techniques such as Artificial Intelligence, Machine Learning, etc., for sentiment analysis, toxicity analysis, and other trends. By leveraging APIs provided by Reddit and 4chan, we implemented a data collection pipeline to collect, process, and store Reddit posts and comments and also 4chan's posts related to the subreddits and board focussed on politics. The collected data provides insights into the behavior of these digital communities. We believe the datasets we have gathered are especially significant since they were collected during the peak election period. Through comparative analysis, we explore the different ideological perspectives and engagement levels across Reddit and 4chan, highlighting the unique discourse patterns of these platforms.

The study captures the vibrant and polarized nature of online conversations revolving around the election, visualized with the help of various data visualization techniques that make use of box plots, bar charts, histograms, word cloud, and more. The datasets section focuses on the different datasets we have gathered, their description and the plots that we have created. Our findings reveal how "Redditors" (Reddit users) and 4chan users react to pivotal moments, including the announcement of Donald Trump as the 47th president of the United States on November 6, 2024. This research not only sheds light on the evolving dynamics of political discussions online but also outlines the challenges of handling large-scale, real-time data for meaningful analysis.

In the conclusion section, we mention the three research questions we have developed based on the gathered data.

Keywords

Reddit API, Crawler, 4Chan API, PostgreSQL, Data Collection, Data Pipeline, Data Measurement and Analysis, Plotting, Visualization, Election Trends, Impact of Social Media

1 Introduction

On November 6, 2024, at 5:34 AM EST, the Associated Press called the US 2024 presidential race and declared Donald Trump as the winner. The 2024 U.S. presidential election marks a pivotal moment in internet history, as online political discourse (podcasts, platforms like X) played a key role in influencing the voters. Platforms such as Reddit and 4chan offer unique ecosystems for political conversations, representing a spectrum of ideologies and engagement styles. This study aims to analyze election-related discussions on these platforms, focusing on sentiment trends, toxicity, and user engagement. Leveraging API-based data collection and NLP tools, our

goal is to provide insights into how different online communities operate.

The motivation behind our work: throughout human history, scholarly debate has been essential in political discourse. Debating various issues while being respectful of the political opponent has always been the ideal way to resolve issues. However, in the current modern era, the anonymity granted by the internet has had some ramifications. It has been easier than ever to abuse others and get away with it due to being anonymous and thousands of miles apart. While harmless trolling and banter will always be a part of online communities, there are some major red flags which should not be ignored. Racist, bigoted remarks made by fringe online groups aimed at minority and marginalized communities are a huge concern. Consistently targeting a group of people with hate speech and hurtful remarks online has a chance of leading to hate crimes in the real world. We believe that identifying and analyzing toxic patterns can help us in deradicalization efforts to ensure a safe and healthy environment for people from all walks of life.

2 Background and related work

Social media has revolutionized political engagement by facilitating real-time discussions and offering valuable insights into public sentiment. Previous studies have emphasized Reddit's role in enabling conversations that often mirror societal divides. Likewise, 4chan's /pol/ board has garnered attention for its high levels of engagement, albeit with frequently controversial content. Building on this body of work, our research focuses on comparing sentiment and toxicity across these platforms, leveraging tools like the Moderate Hate Speech API and SpaCy to analyze emotional and ideological disparities.

3 Table

Metrics	Reddit posts	Reddit comments	4chan posts
Total volume	6846	955,874	1,828,989
Total volume in r/politics	5307	915,675	N/A
Average comments per post	N/A	286	N/A
Peak Volume	686 (6 th Nov. 2024)	210,228 (6 th Nov. 2024)	240,619 (6 th Nov. 2024)
Avg post/day	480	N/A	130,642
Avg comments/day	N/A	69,930	N/A

Figure 1: Metrics Comparison for Reddit Post, Reddit Comments and 4Chan Threads

4 Dataset

- Data Sources:
 - 4chan: A total of 1,828,989 posts were collected over 14 days, providing unfiltered and often extreme discourse

on political topics. The word cloud (Figure 1) highlights the prominence of terms like "Trump," "economy," and "tariffs," reflecting discussions centered on the 2024 U.S. presidential election and policy issues.

- Reddit Comments: Collected data includes 955,874 comments from subreddits like r/politics, r/neutral, etc, showcasing structured discourse focusing on the election, candidates, and key policy matters. The word cloud and sentiment analysis timeline emphasize the focus on key political events and the polarized nature of discussions.
- Reddit Posts: Over 6846 posts were analyzed, reflecting overarching narratives on policies and campaign-related topics. The word cloud and frequency analysis of policy keywords show how Reddit posts highlight policy discussions and broader electoral themes.
- Dataset Comparisons:
 - Word clouds reveal shared themes (e.g., "Trump," "vote," "election") but highlight platform-specific discourse patterns. While 4chan exhibits a mix of provocative and policy-related keywords, Reddit data reflects structured discussions centered on policy issues and election-related events.
 - Frequency charts show keyword prominence differences: terms like "economy" and "inflation" dominate for Trump-related mentions, while Harris-related mentions focus on topics like "equal pay" and "abortion."
- Sentiment Analysis:
 - The sentiment timeline highlights distinct peaks in positive sentiment (e.g., during election results on November 6) and negative dips, showcasing polarized reactions to political events. These fluctuations align with key moments in the 2024 U.S. election cycle.

5 Plots

- In Figure 2, this stacked bar chart represents user engagement (likes and comments) on Reddit posts surrounding key political events from November 1 to November 14, 2024. Significant spikes are visible on November 5, 6, and 7, corresponding to pivotal dates during the U.S. election. Engagement is predominantly driven by likes, with comments adding a smaller but notable portion. The chart highlights the surge in activity during major political events.
- In Figure 3, this bar chart compares the frequency of mentions for various policy keywords in comments related to candidates Trump and Harris. Trump dominates discussions on "economy," "inflation," "border," and "tariffs," while Harris shows higher mentions for topics like "equal pay" and "abortion." Both candidates share mentions on keywords such as "immigration" and "security," although these topics show a stronger association with Trump. The chart captures the distinct policy narratives shaping discussions around each candidate, shedding light on their perceived priorities and public resonance.
- In Figure 4, this timeline chart shows the hourly sentiment analysis of comments on Reddit's r/politics subreddit between November 1 and November 14, 2024. Positive sentiment (gold) and negative sentiment (purple) scores fluctuate significantly, with spikes around key political events. A major positive peak is observed on November 6, corresponding to the U.S. election results, while a pronounced negative dip occurs immediately after, reflecting polarized reactions.
- In Figure 5, this horizontal bar chart displays the top 30 most frequent words associated with Trump in discussions. Prominent terms like "vote," "election," "voters," and "president" underscore a focus on the 2024 U.S. election and governance.
- In Figure 6, the horizontal bar chart displays the top 30 most frequent words associated with Harris in comments on the r/politics subreddit during the 2024 election. Prominent terms like "Trump," "vote," "Biden," and "election" underscore the interconnected discussions about candidates and the election process. Other terms, such as "campaign," "Democrats," and "win," highlight the political and policy-focused nature of the discourse. Words like "women," "states," and "better" suggest engagement with broader societal and policy issues. The chart illustrates the election's central themes and Harris's role in political discussions.
- In Figure 7, the bar chart compares the frequency of policy keyword mentions for two U.S. presidential candidates, Trump and Harris, on 4chan. Key insights include a significantly higher number of mentions for Harris compared to Trump for keywords like "economy," "equal pay," and "care," while Trump has slightly higher mentions for "immigration" and "national security." The distribution of mentions suggests differing areas of focus or controversy for each candidate within the platform's discussions.
- In Figure 8, this word cloud visualizes the most frequently used words from 4chan discussions. Prominent words include profane language, offensive slurs, and political keywords, reflecting the platform's often aggressive and polarizing discourse. The size of each word correlates with its frequency in the dataset, providing insight into the dominant themes and language patterns used.
- In Figure 9, this world map visualizes sentiment analysis of 4chan discussions by country, offering insights into global reactions to the 2024 U.S. presidential election. The sentiment scale ranges from negative (red) to positive (blue). Key observations include:
 - Negative Sentiment: Countries like Greenland and parts of Eastern Europe show strong negative sentiment, possibly indicating disapproval of election outcomes or the political climate in the U.S.
 - Positive Sentiment: Select African nations, such as Ghana and Kenya, exhibit higher positive sentiment, suggesting optimism or approval, potentially influenced by U.S. foreign policies or candidate narratives resonating in these regions.
 - Neutral to Mildly Negative Sentiment: North and South America, along with parts of Asia and Australia, largely reflect neutral to mildly negative sentiment, hinting at either political neutrality or moderate discontent.

The map highlights how regional issues, cultural perceptions, and the global impact of U.S. politics shape the tone of discussions worldwide. It underscores the varied emotional responses driven by the 2024 election and the influence of different geopolitical contexts

- In Figure 10, this word cloud highlights the key terms and sentiment across 4chan threads, Reddit comments, and Reddit posts during the 2024 election. Across all datasets, terms like "Trump," "election," and "vote" are central. However 4chan Threads: Combine election-related terms with hateful language, reflecting unmoderated, provocative discourse. Reddit Comments and Posts: Focus on election discussions and candidates, with less offensive language, showcasing a more moderated environment.
- In Figure 11, for comments classified normal (non-toxic): The median is extremely close to 1.0, indicating that Moderate Hate Speech API is highly confident in labeling comments as non-toxic. The IQR is almost non-existent, and the whiskers are very short. This means that confidence scores for non-toxic comments are highly clustered near 1.0, with little variability. We concluded that non-toxic classifications are very consistent and reliable. For comments classified flag (toxic): The median is approximately 0.85, which is significantly lower than that of non-toxic comments. The IQR is much larger than for non-toxic comments. Confidence scores for toxic comments are spread across a wider range, indicating more inconsistency in toxic classifications. The whiskers extend farther compared to non-toxic comments, showing Moderate Hate Speech API's uncertainty and variability in assigning scores for toxic comments.
- In Figure 12, the stacked bar chart shows the 55,542 comments that were generated in r/politics in the time period between 6 November, 2024 5:30 AM and 6 November, 2024 2:30 PM. Since the AP called the race in favor of Donald Trump on 6 November, 2024 5:30 AM, these are the comments that were generated immediately after that. We processed these comments using the Moderate Hate Speech API and 53,726 of those comments were classified as normal, while 1816 comments were classified as flag.
- In Figure 13, the bar chart shows the number of submissions made daily on the r/politics subreddit from November 1 to November 14, 2024. A clear peak is visible on November 6, coinciding with the day of the U.S. presidential election when Donald Trump won.
- In Figure 14, the chart visualizes the number of comments posted per hour on the r/politics subreddit over the first two weeks of November 2024. A significant spike is observed on November 6, indicating a surge in activity likely tied to a major political event. The rest of the data shows daily patterns of engagement with peaks and troughs.
- In Figure 15, the plot above shows the number of comments per hour on 4chan's /pol/ board from November 1st to November 14th, 2024. The biggest spike occurred on November 6th, 2024, the day of the Trump election win, reflecting a surge in user engagement and activity on the platform in response to this significant political event.

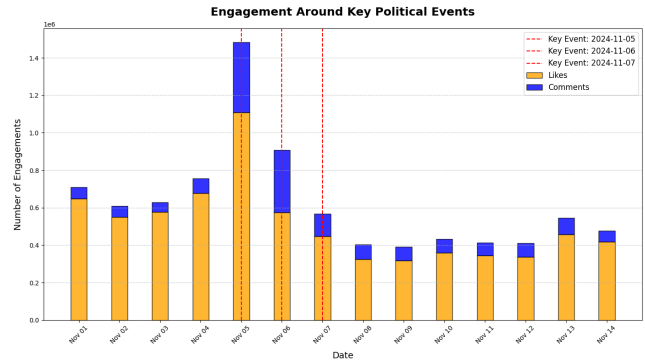


Figure 2: Engagement Metrics Around Key Political Events on Reddit During the 2024 U.S. Election (November 1–14, 2024)

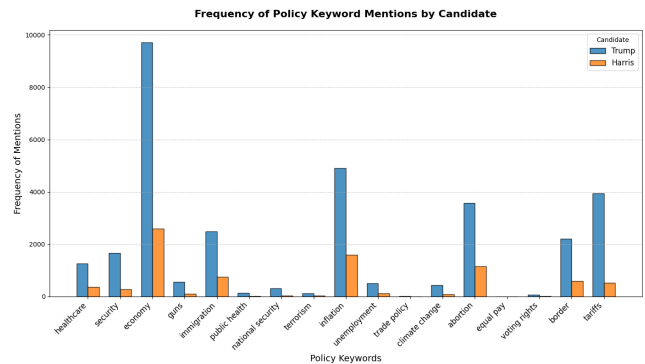


Figure 3: Frequency of Policy Keyword Mentions in Comments by Candidate on r/politics Subreddit (2024 Election)

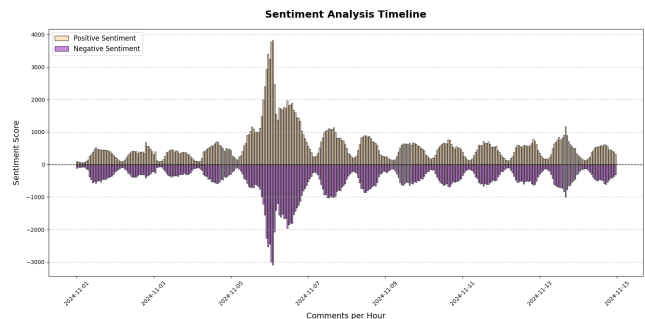


Figure 4: Hourly Sentiment Analysis of Comments on r/politics Subreddit (Nov 1–14, 2024)

6 Discussion

Across all platforms, the discourse is heavily focused on the 2024 U.S. presidential election, with keywords like "Trump," "Harris," and "vote" dominating discussions (Figures 10).

- Platform-Specific Differences:

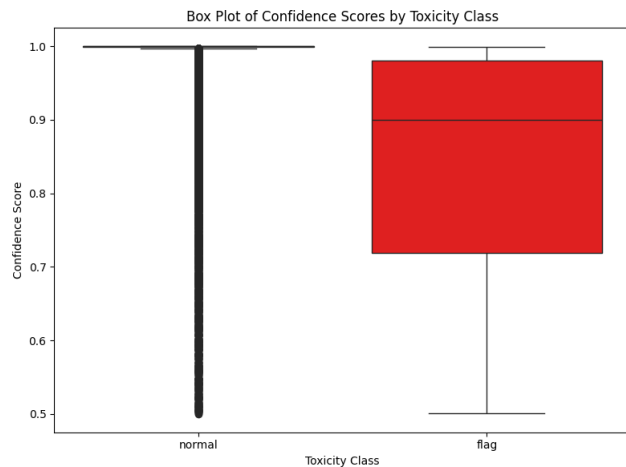


Figure 11: Box Plot of confidence score by Toxicity Class

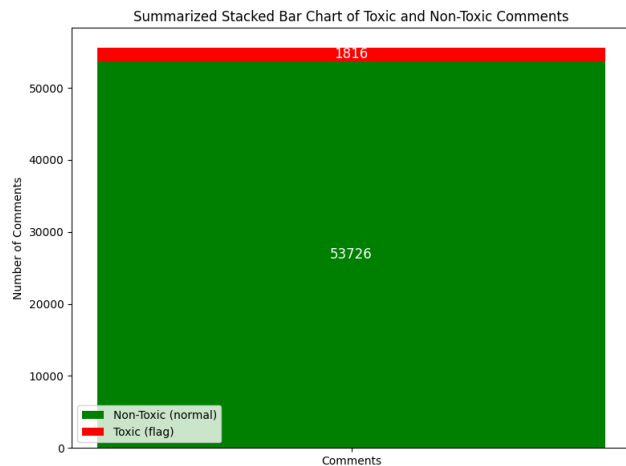


Figure 12: Stacked Bar Chart of Toxic and Non-Toxic Comments

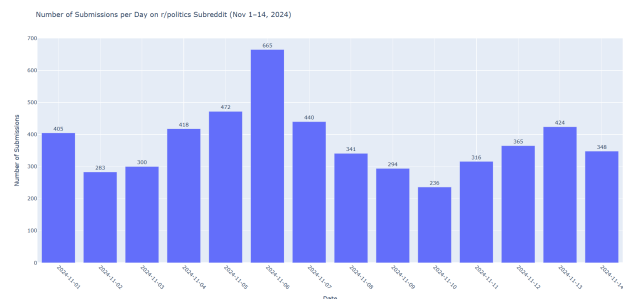


Figure 13: Daily Submissions on r/politics Subreddit During the 2024 U.S. Election (Nov 1-14, 2024)

- Research Question 1: Analyzing toxicity and sentiments across different subreddits and 4chan's /pol board.

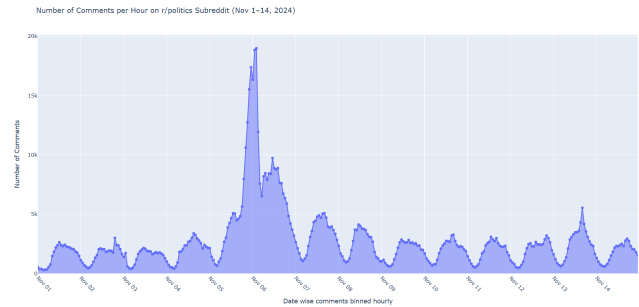


Figure 14: "Hourly Comment Activity on r/politics Subreddit (Nov 1-14, 2024)"

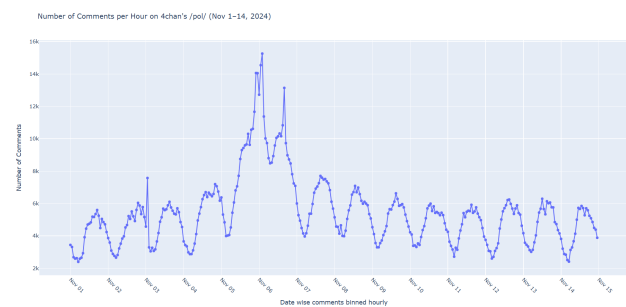


Figure 15: Hourly comment activity on 4chan's /pol/ board

- Research Question 2: Using Named entity recognition (NER) or topic modeling to identify entities or topics being talked about.
- Research Question 3: How does platform-specific language and tone influence public perception of that particular platform and the issues being discussed therein?

7 Conclusion, Limitations, and Future Work

This study highlights how political discussions play out on platforms like Reddit and 4chan during a major election. By analyzing sentiment, toxicity, and engagement, we've shown the unique traits and ideological leanings of these communities:

- Moderation matters: Reddit fosters more balanced discussions, while 4chan amplifies unfiltered, often extreme opinions.
- Campaigns can use these insights to better target their messages based on platform-specific trends and user behavior.
- Limitations:
 - We focused only on text data, missing insights from images and videos.
 - The analysis was limited to a few subreddits and 4chan's /pol/ board, which may not capture the full picture.
 - The short-lived nature of 4chan content made data collection tricky, leaving some gaps.
 - The Moderate Hate Speech API used was sometimes slow and had outages, causing a delay in data collection.

- The [/r/subreddit]/comments/article endpoint doesn't give all the comments under a post. It only gives max 500 comments regardless of the value of the limit parameter. Even using the /api/morechildren endpoint after that doesn't give all the remaining comments, only some of them.
- Future Work:
 - Adding more platforms like Youtube could give a broader perspective.
 - Including multimedia content like images and audio could improve the depth of analysis.
 - Building NLP tools tailored to the language styles of each platform could refine sentiment and toxicity detection.
 - Making the system robust to handle periodic outages and slowness from APIs like Moderate Hatespeech API.

8 Citations

We have referenced the following articles to guide the implementation of our project and the preparation of this report:

- Kek, Cucks, and God Emperor Trump: A Measurement Study of 4chan's Politically Incorrect Forum and Its Effects on the Web [1]
- The Reddit Politosphere: A Large-Scale Text and Network Resource of Online Political Discourse[2]
- iDRAMA-Scored-2024: A Dataset of the Scored Social Media Platform from 2020 to 2023 [3]

References

- [1] Gabriel Emile Hine, Jeremiah Onalapo, Emiliano De Cristofaro, Nicolas Kourtellis, Ilias Leontiadis, Riginos Samaras, Gianluca Stringhini, and Jeremy Blackburn. 2017. Kek, Cucks, and God Emperor Trump: A Measurement Study of 4chan's Politically Incorrect Forum and Its Effects on the Web. *arxiv.org* (Oct. 2017). <https://arxiv.org/pdf/1610.03452>
- [2] Valentin Hofmann, Hinrich Schutze, and Janet B. Pierrehumbert. 2022. The Reddit Politosphere: A Large-Scale Text and Network Resource of Online Political Discourse. *Association for the Advancement of Artificial Intelligence* (2022). <https://epub.ub.uni-muenchen.de/107434/1/19377-Article%20Text-23390-1-2-20220531.pdf>
- [3] Emiliano De Cristofaro Gianluca Stringhini Jeremy Blackburn Jay Patel, Pujan Paudel. 2024. iDRAMA-Scored-2024: A Dataset of the Scored Social Media Platform from 2020 to 2023. *arxiv.org*, Article 2405.10233 (2024). <https://arxiv.org/html/2405.10233v1>