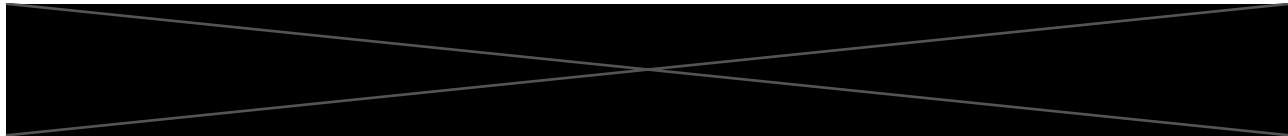


Analyzing North American Media Coverage in North America on President Donald Trump



Abstract

Media coverage has a strong influence on public perception of political figures. This report analyzes North American media coverage of President Trump across 694 news articles. Using an open coding framework developed from 200 initial articles, we identified six thematic categories and assessed sentiment orientation (positive, negative, or neutral) across the collected articles. TF-IDF analysis validated our coding framework by revealing highly relevant terms within each topic cluster. Additionally, our findings suggest that, contrary to common perceptions about media bias, most articles maintained neutral tone and objective reporting.

Introduction

We considered several political figures for the media analysis, but we ultimately selected President Trump as our choice due to practical research considerations and opportunities. President Trump is a key figure in American and International politics - as a former president, current president-elect, and a figure at the center of several legal hearings.

These characteristics of President Trump have allowed for an abundance of coverage to be accessible through NewsAPI. Additionally, his involvement in multiple significant events simultaneously (legal proceedings, election campaign, election results reporting, and others) have offered an exciting opportunity to examine how media outlets prioritize and combine different aspects of their coverage on a figure.

Our research questions focused on two key dimensions:

- What topics dominate Trump coverage in recent news publications, and what are the most distinctive terms appearing in the coverage?
- How does sentiment vary in these publications?

To mitigate and minimize biases, we drew articles from various news outlets across the political spectrum. This sampling allowed us to identify patterns in how different outlets approach Trump-related coverage - moving beyond partisan divisions to uncover deeper trends.

As a result, we refined our research questions to the following:

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

- What topics dominate Trump coverage in North American news articles published in the last 28 days, and what are the most distinctive terms appearing in the coverage for each topic?
- How does sentiment vary within different topic areas of coverage?

Data

Our dataset captures detailed information on 694 articles collected to analyze media coverage of President-elect Donald Trump's activities, focusing on key topics such as Cabinet appointments, policy decisions, and public and political reactions. The structured dataset contains the following columns:

The Topic and Sentiment columns were developed through an open coding framework

- **Source ID:** A unique identifier representing the source (e.g., "abc-news").
- **Source Name:** The name of the news organization (e.g., "ABC News").
- **Author:** The author(s) of the article, if available.
- **Title:** The headline of the article.
- **Description:** The subtitle of the article, or if unavailable, the first sentence of the body.
- **URL:** A link to the full article.
- **URL to Image:** A link to an image associated with the article.
- **Published Date:** The publication date and time of the article.
- **Content:** The main text of the article, truncated when exceeding the API's character limit.
- **Topic:** Human-annotated category of the article's content (e.g., "Policy Proposals and Political Positions," "Elections Results and Coverage").
- **Sentiment:** Human-annotated tone of the article, categorized as Positive, Negative, or Neutral.

This structure provides a comprehensive foundation for analyzing trends and framing in media coverage.

Data Collection Process

The dataset was constructed using **NewsAPI**, accessed through a custom Python script. The script utilized the ‘everything’ endpoint with targeted keywords and filtering parameters to ensure the relevance and quality of results.

- **Keywords and Filters:** Queries incorporated specific terms such as “Trump,” “Cabinet,” and “policy” to capture relevant articles. Language was restricted to **English**, ensuring linguistic consistency. The collection period spanned **October 20 - November 17, 2024**, focusing on both pre-election and immediate media reactions post-election.
- **API Limitations:** The free plan of NewsAPI restricts data collection to a 28-day window. To overcome this, the script issued daily queries for articles within the specified timeframe. Each query was capped at 100 results, necessitating iterative requests over individual days to ensure coverage.
- **Source Selection:** While the API indexed many credible sources, some prominent outlets were inaccessible due to API restrictions. To compensate, a diverse range of available sources was included to maximize representativeness.

Table 1: Distribution of News Sources and Their Political Orientation (AllSides 2024)

News Source	Articles	Political Orientation
ABC News	194	Center Left
CBS News	190	Center
Fox News	113	Center-Right
NBC News	78	Center-Left
Business Insider	48	Center-Left
CBC News	32	Center-Left
CNN	23	Center-Left
Associated Press	10	Center
USA Today	5	Center
Bloomberg	1	Center-Left
Total	694	

*Based on AllSides Media Bias Ratings (2024)

Bias Avoidance

Sampling was conducted with deliberate measures to minimize biases and ensure representativeness. Articles were collected from a diverse range of sources indexed by NewsAPI, covering outlets across the political spectrum to balance ideological perspectives. Temporal alignment was maintained by restricting collection to a uniform three-day timeframe, ensuring comparability of media coverage during the same political events. To avoid linguistic or cultural inconsistencies, only English-language articles were included. Additionally, targeted keyword refinement ensured relevance, capturing articles explicitly related to Trump while filtering out tangential or unrelated content. These measures collectively minimized sampling bias, ensuring a robust and balanced dataset.

Data Cleaning and Filtering

After collection, the dataset underwent a thorough cleaning process to ensure quality and relevance:

- **Duplicate Removal:** Articles with identical titles, URLs, or content were removed.
- **Relevance Filtering:** Articles were filtered to retain only those explicitly discussing Trump-related topics, ensuring alignment with the research focus.
- **Language Consistency:** Non-English articles were excluded.
- **Missing Data:** Entries with incomplete critical fields (e.g., title or description) were removed.

Despite these measures, certain challenges persisted:

- The “content” field was often truncated by the API, limiting the completeness of data for some entries.
- The API’s “relevancy” indexing occasionally returned irrelevant articles, which were subsequently excluded through manual review.

Challenges and Adjustments

The following challenges were encountered during data collection and preparation:

- **Lookback Limitations:** The API’s 30-day limit constrained the timeframe of analysis. This was addressed by focusing on recent events and iteratively querying daily data within the allowed window.
- **Result Caps:** With a maximum of 100 results per query, the script implemented a granular daily collection process, aggregating results over multiple iterations to achieve sufficient sample size.
- **Source Restrictions:** Several high-traffic outlets were unavailable, impacting the breadth of coverage. However, the inclusion of a diverse array of accessible sources helped mitigate this issue.
- **Keyword Ambiguity:** Despite targeted keywords, some articles were tangentially related or irrelevant. Post-collection filters ensured that “Trump” appeared in either the title or description.
- **Data Truncation:** The truncation of the “content” field required reliance on other attributes, such as title and description, for accurate categorization and sentiment analysis.

By implementing these steps, our analysis ensured that the resulting TF-IDF scores accurately reflected the thematic distinctiveness of terms within each topic while addressing the unique characteristics of the dataset.

Final Dataset

The final dataset is a high-quality, annotated collection of 694 articles that provides comprehensive insights into media coverage trends. Its structure and content are designed to support in-depth analysis of framing and thematic representation in media narratives surrounding Trump’s political activities. By addressing API constraints and implementing rigorous cleaning and annotation processes, the dataset achieves a balance of breadth, depth, and analytical utility.

Methods

Typology Development

As part of our analysis effort, we systematically developed a comprehensive and sharply-defined typology consisting of 6 categories. The process began by annotating 33% of the data; each of the group members independently annotated 11% of the data to start. This approach allowed to ensure a robust typology development while reserving sufficient data for subsequent validation.

Splitting the initial phase annotation into the different group members also allowed for different perspectives and insights. Additionally, the independent nature of these initial annotations was critical for minimizing potential biases; each annotator developed their annotations independently without being influenced by others' results.

After the initial annotations were completed, the team held discussions to reconcile the different categories developed and aggregate them into 6 mutually exclusive and collectively exhaustive categories, each with a clear definition and explicit inclusion/exclusion criteria.

Annotation Process

To enhance the dataset's analytical value, human annotation was performed on two key dimensions:

- **Topic Categorization:** Articles were categorized into themes such as "Policy Proposals and Political Positions," "Elections Results and Coverage." These categories were developed iteratively through the above described typology development process.
- **Sentiment Analysis:** Each article was labeled as Positive, Negative, or Neutral based on its tone and framing. Ambiguous cases were resolved through team consensus.

TF-IDF Calculation Process

To extract the most distinctive terms within each topic, we implemented a TF-IDF calculation tailored to our dataset. This process ensured that the results accurately reflected the thematic nuances of the data.

- **Preprocessing the Text Data:** The raw dataset underwent a detailed preprocessing phase. Text from the title and description fields was tokenized into individual words. Stop words—common but uninformative terms—were removed using the `tidytext` package's default stop word list. Additionally, we curated a custom stop word list specific to the dataset, which included words such as "trump," "donald," and "2024." These terms were excluded to minimize noise and enhance the focus on meaningful terms.
- **Term Frequency (TF):** The term frequency for each word was calculated within its respective topic. This metric quantified the frequency of a word's appearance, capturing its importance within the context of the topic.
- **Inverse Document Frequency (IDF):** To account for the distribution of terms across topics, the inverse document frequency was computed. IDF was defined as the logarithm of the ratio of the total number of topics to the number of topics in which the term appeared. This step

downweighted commonly occurring terms across multiple topics.

- **TF-IDF Score Calculation:** The TF and IDF scores were multiplied to derive the TF-IDF score for each word. High TF-IDF scores identified words that were frequent within a specific topic but rare across others, emphasizing their relevance to the topic's unique context.
- **Visualization:** Bar charts were generated to visualize the top terms for each topic based on their TF-IDF scores. The visualizations included additional customizations, such as standardized bar colors and numeric TF-IDF values displayed on the bars, to enhance interpretability. These bar charts provided a clear representation of the key terms that defined each topic.
- **Challenges and Adjustments:** Several challenges were addressed during the TF-IDF calculation process. Noise introduced by high-frequency, dataset-specific terms was mitigated through the use of custom stop words. The variability in text length between titles and descriptions was balanced by combining the fields into a unified text column before processing. These adjustments ensured that the TF-IDF scores provided a robust and accurate depiction of the thematic distinctiveness within the dataset.

Results

Motivation & Context of Topology Development

The development of this typology is essential for several reasons:

- To systematically analyze the a wide array of Trump-related media coverage
- To enable consistent classification of news articles for our analysis purposes
- To identify patterns in media coverage across different aspects of Trump's political and social activities

We developed a typology of six distinct, exhaustive, and comprehensive categories, described below.

1. Policy Proposals and Political Positions (n = 134):

Definition Articles that primarily focus on specific policy initiatives, legislative proposals, executive actions, or clearly stated political positions by Trump or his associates.

Positive Example

- "What Trump's mass deportation plan might look like"
 - Included because it:
 - Centers on a specific policy proposal
 - Analyzes implementation feasibility
 - Examines potential impacts

Negative Example

- "Trump rallies supporters in Florida" - Excluded because it:
 - Focuses on campaign activities rather than policy

Edge Cases

- Articles discussing policy implications during campaign speeches:
 - Classified based on the depth of policy discussion

2. Elections Results and Coverage (n = 118):

Definition Articles focusing on electoral outcomes, voting patterns, post-election polling analysis, and media coverage of election-related events and trends.

Positive Example

- "Fact check: Watch out for these election night falsehoods" - Included because it:
 - Directly addresses electoral information
 - Analyzes election-specific claims

Negative Example

- "Trump's economic vision for second term" - Excluded because it:
 - Focuses on policy rather than electoral process
 - Lacks significant electoral analysis

Edge Cases

- Articles combining polling data with policy analysis:
 - Classified based on primary focus
 - Included if electoral analysis dominates

3. Reaction and Public Sentiment (n = 145):

Definition Captures societal, emotional, or grassroots reactions to political events, statements, or decisions, including public discourse in forums and media.

Positive Example

- "Americans in Puerto Rico can't vote for US president. Their anger at Trump is shaping the race" - Included because it:
 - Highlights public frustration and debate over voting rights.
 - Explores grassroots sentiment and calls for electoral reform.
 - Focuses on societal responses to systemic inequities.

Negative Example

- "Trump still hasn't signed ethics agreement required for presidential transition" - Excluded because it:
 - Focuses on a legal or procedural issue rather than societal or grassroots reactions.
 - Lacks emotional or public discourse elements central to this category.

Edge Cases

- Articles discussing societal reactions to specific policies:
 - Classified based on whether the focus is on the reaction itself or the policy details.

4. Legal Issues and Controversies (n = 85):

Definition Articles on legal cases, criminal allegations, or ethical misconduct involving political figures or institutions.

Positive Example

- "Judge sets hearing on \$1M-a-day sweepstakes from Elon Musk PAC helping Donald Trump" - Included because it:
 - Focuses on a case involving financial manipulation.
 - Discusses court proceedings and campaign funding implications.
 - Highlights legal scrutiny of political figures.

Negative Example

- "SNL skewers Trump's picks for cabinet positions" - Excluded because it:
 - Focuses on satire and entertainment, not legal controversies.
 - Lacks legal substance or discussion of ethical scrutiny.

Edge Cases

- Articles involving ethical debates during campaigns:
 - Classified based on whether the article centers on legal proceedings or campaign strategies.

5. Foreign Affairs and Relations (n = 44):

Definition Articles focusing on global diplomacy, international conflicts, foreign aid, or U.S. involvement in foreign affairs, with analysis and examples.

Positive Example

- "Trump gives Netanyahu deadline to resolve the Israel-Hamas conflict" - Included because it:
 - Directly addresses Trump's involvement in a major foreign policy issue
 - Analyzes diplomatic actions and their potential global impact
 - Focuses on a high-stakes international conflict driven by U.S. political decisions

Negative Example

- "Trump's tariffs shake global markets" - Excluded because it:
 - Focuses primarily on economic outcomes rather than diplomacy or international relations
 - Lacks analysis of specific foreign policy actions or diplomatic context
 - Discusses global market dynamics without direct links to international conflict or foreign aid

Edge Cases

- Articles combining foreign policy with economic impacts:
 - Classified based on the dominant focus
 - Included if Trump's diplomatic efforts are central

6. Electoral Campaign and Strategies (n = 168):

Definition Articles focusing on candidate strategies, public events, voter outreach, pre-election polling data, and campaign dynamics with examples and analysis.

Positive Example

- "Trump intensifies rallies in key battleground states ahead of 2024 election" - Included because it:
 - Focuses on Trump's strategic voter outreach in critical regions
 - Analyzes campaign efforts to secure electoral advantages
 - Highlights the importance of battleground states in Trump's re-election bid

Negative Example

- "Trump's vision for his next term focuses on tax cuts"
 - Excluded because it:
 - Primarily discusses policy proposals rather than campaign strategy
 - Lacks significant analysis of voter engagement or electoral dynamics
 - Centers on governance plans instead of election-related efforts

Edge Cases

- Articles combining campaign strategy with policy discussion:
 - Classified based on the dominant focus
 - Included if voter outreach or electoral strategy is central

Evidence for Comprehensiveness

The comprehensiveness of these categories is supported by:

- TF-IDF analysis showing distinct vocabulary clusters and minimal overlap in key terms
- The categories being mutually exclusive and collectively exhaustive

Sentiment Coding

The sentiment analysis methodology follows specific criteria to ensure consistent classification across all articles. Articles are categorized as negative when they predominantly feature critical language, reports of failures or controversies, direct criticism from credible sources, fact-checked contradictions of Trump's statements, or descriptions of negative consequences of his actions or policies. Neutral articles maintain objective distance by presenting multiple balanced viewpoints. They focus on factual reporting, using primarily statistical data and direct quotes without evaluative language, or covering routine political processes without editorial commentary. Positive sentiment is assigned to articles that highlight achievements, successful policy implementations, favorable economic and social indicators during his presidency, supportive statements from allies or experts, or positive public reception of his initiatives.

Discussion

Bar charts were used to showcase the top terms for each topic, emphasizing the distinguishing words that define the thematic content. The following bar charts provide a clear representation of the most influential terms within each category.

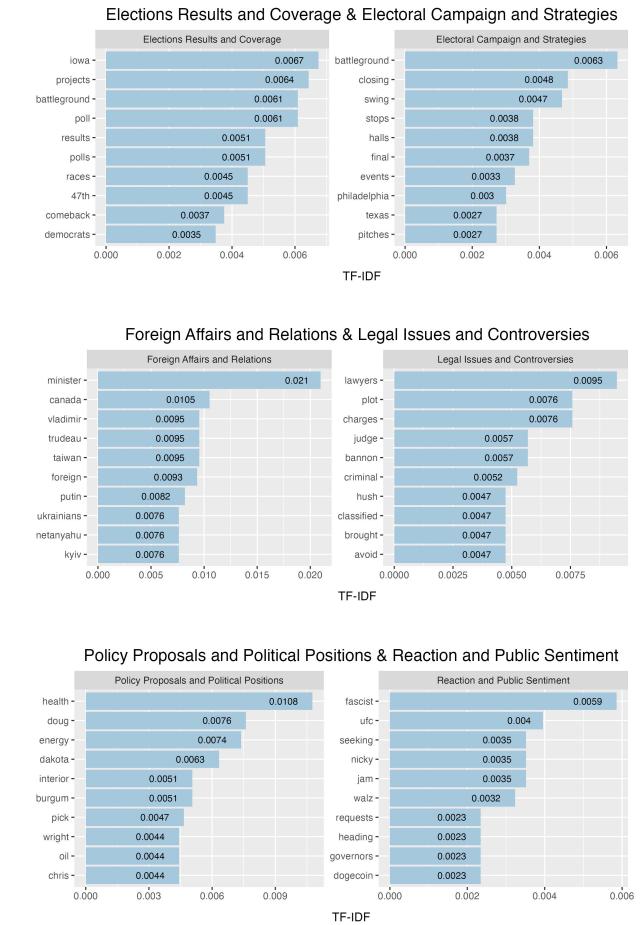


Figure 1: Top TF-IDF terms and their corresponding scores across six thematic categories in Trump-related media coverage, illustrating the most distinctive vocabulary associated with each topic category.

TF-IDF Results Analysis

The TF-IDF analysis shows insightful patterns in how media outlets covered Trump-related activity in the past month.

- 1. Elections Results and Coverage** In the "Elections Results and Coverage" Category, "Iowa" emerges with the highest TF-IDF (0.0067); this reflects Iowa's importance in primary election coverage. As Prokop (Prokop 2023) explains, "It's because Iowa is the first state where actual voters across any US state get up and go say who they want to be president." Following this are terms like "projects," "poll," and "results," which highlight the me-

dia's focus on projecting and analyzing election outcomes. Furthermore, terms like "47th" and "comeback" reflect coverage surrounding Trump's campaign narrative about potentially becoming the 47th President of the United States.

2. **Electoral Campaign and Strategies** In the "Electoral Campaign and Strategies" category, "battleground" reappears with a score of 0.0063, which shows its relevance not only to election results but also to Trump's campaign strategies. Additional terms like "closing" and "swing" (0.0048 and 0.0047, respectively) point to media focus on swing states and their potential impact on the campaign. Meanwhile, geographic terms such as "Philadelphia" and "Texas" (0.0030 and 0.0027) indicate less distinctive coverage of state-wide campaign events. Similarly, campaign-specific terms like "stops" and "halls" have moderate TF-IDF scores, suggesting that while these activities were covered, they did not dominate media coverage.
3. **Foreign Affairs and Relations** The TF-IDF analysis of "Foreign Affairs and Relations" highlights significant media attention on geopolitical topics. The term "minister" scores the highest at 0.021, reflecting a focus on diplomatic and governmental figures. Following closely are terms such as "canada," "vladimir," and "trudeau" (0.0105, 0.0095, and 0.0095 respectively), emphasizing the media's focus on specific international figures and bilateral relations. Additionally, terms like "taiwan" and "foreign" (0.0095 and 0.0093) suggest attention to global tensions and foreign policies, while "putin," "ukrainers," and "kyiv" (0.0082, 0.0076, and 0.0076 respectively) point to coverage of the ongoing conflict in Ukraine.
4. **Legal Issues and Controversies** In the "Legal Issues and Controversies" category, the term "lawyers" has the highest score of 0.0095, highlighting the media's focus on Trump's legal battles. Terms like "plot" and "charges" (both scoring 0.0076) reflect the media's coverage of legal allegations and criminal investigations tied to Trump and his associates. Similarly, "judge," "bannon," and "criminal" (0.0057, 0.0057, and 0.0052 respectively) suggest a focus on courtroom proceedings and high-profile cases involving individuals in Trump's circle. Meanwhile, terms like "classified" and "hush" (0.0047 each) allude to coverage of controversies surrounding classified documents and alleged hush-money payments, though these topics received comparatively less emphasis.
5. **Policy Proposals and Political Positions** The TF-IDF analysis of "Policy Proposals and Political Positions" highlights media focus on specific policy areas tied to President Trump's political stance. The term "health" has the highest TF-IDF score of 0.0108, indicating significant attention to healthcare policies with the nomination of Robert F. Kennedy Jr. to serve as the United States Secretary of Health and Human Services (Reuters 2024). Following this are terms like "doug" and "energy" (0.0076 and 0.0074, respectively), pointing to cov-

erage of prominent figures or discussions about energy policies under Trump's administration, with Doug referencing Ontario Premier Doug Ford. The term "dakota" (0.0063) suggests a connection to issues like the Dakota Access Pipeline, which was a topic of environmental and economic focus during his presidency. Lower-scoring terms like "interior," "oil," and "chris" (0.0051, 0.0044, and 0.0044) reflect broader, but less emphasized, policy coverage.

6. **Reaction and Public Sentiment** In the "Reaction and Public Sentiment" category, the term "fascist" has the highest TF-IDF score of 0.0059, underscoring polarizing public reactions and strong language in media narratives about Trump. Terms such as "ufc" (0.004) reflect social contexts in the media coverage of Trump-related events, such as President Trump attending a UFC event in Madison Square Garden alongside Elon Musk and other associates (Weissert and Gelston 2024). Similarly, "nicky jam" and "dogecoin" (both scoring 0.0035 and 0.0023, respectively) indicate a mix of public sentiment and cultural references. Nicky Jam is a Puerto Rican artist who withdrew his support of President Trump in the beginning of November - an event that was highly covered in the media. Additionally, DOGE refers to both Dogecoin's sharp increase in price following Trump's win, in addition to Trump creating the Department of Government Efficiency (DOGE), which was an event highly covered in the media.

Sentiment Analysis

The heatmap and pie charts together offer complementary perspectives on the distribution of sentiments—negative, neutral, and positive—across thematic categories. The heatmap provides a high-level view of sentiment prevalence, highlighting which sentiments dominate specific categories. In contrast, the pie charts shift the focus to the categories themselves, breaking down the sentiment distribution within each. This dual approach ensures a nuanced understanding of both overarching trends and the detailed sentiment dynamics of individual categories, setting the stage for the deeper analyses that follow.

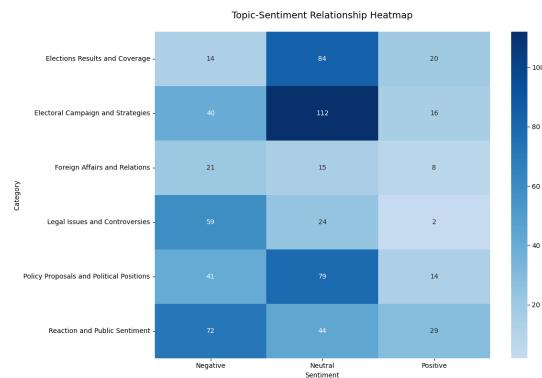


Figure 2: Heatmap describing the relationship between each thematic category and the three sentiment levels

Thematic Category Distribution across Sentiment

Negative Sentiment: Negative sentiment is prominent in contentious or emotional topics:

- "Reaction and Public Sentiment": 72 articles, highlighting public discontent.
- "Legal Issues and Controversies": 59 articles, reflecting critical reporting on scandals and legal disputes.
- "Policy Proposals and Political Positions": 41 articles, focusing on policy critiques.
- "Electoral Campaign and Strategies": 40 articles, tied to campaign challenges.
- "Foreign Affairs and Relations": 21 articles, reflecting diplomatic tensions.
- "Elections Results and Coverage": Only 14 articles, indicating a more neutral tone.

Neutral Sentiment: Neutral sentiment dominates most categories, reflecting balanced reporting:

- "Electoral Campaign and Strategies": 112 articles, emphasizing balanced reporting.
- "Elections Results and Coverage": 84 articles, focusing on objective reporting of election outcomes.
- "Policy Proposals and Political Positions": 79 articles, reflecting factual analysis.
- "Reaction and Public Sentiment": 44 articles, presenting a neutral observations on public responses.
- "Legal Issues and Controversies": 24 articles, balancing critical narratives with factual reporting.
- "Foreign Affairs and Relations": Only 15 articles, showing more polarized reporting.

Positive Sentiment: Positive sentiment, though less frequent, highlights moments of success:

- "Reaction and Public Sentiment": 29 articles, capturing moments of public approval.
- "Elections Results and Coverage": 20 articles, recognizing electoral success.
- "Electoral Campaign and Strategies": 16 articles, emphasizing effective strategies.
- "Policy Proposals and Political Positions": 14 articles, recognizing well-received policies.
- "Foreign Affairs and Relations": 8 articles, focusing on positive diplomatic developments.
- "Legal Issues and Controversies": Only 2 articles, reflecting its challenging nature.

Sentiment Distribution Across Categories

- **Elections Results and Coverage:** Neutral sentiment dominates (71.2%), with smaller proportions of negative (16.9%) and positive (11.9%).
- **Electoral Campaign and Strategies:** Neutral sentiment is prevalent (66.7%), followed by negative (23.8%) and positive (9.5%).

- **Reaction and Public Sentiment:** A more diverse distribution is observed, with neutral (49.7%), negative (30.3%), and positive (20%).
- **Policy Proposals and Political Positions:** Neutral sentiment (59%) is the largest proportion, while negative (30.6%) and positive (10.4%) are less frequent.
- **Legal Issues and Controversies:** Negative sentiment dominates (69.4%), with smaller shares of neutral (28.2%) and positive (2.4%).
- **Foreign Affairs and Relations:** Neutral sentiment (47.7%) is most prominent, with negative (34.1%) and positive (18.2%) sentiments following.

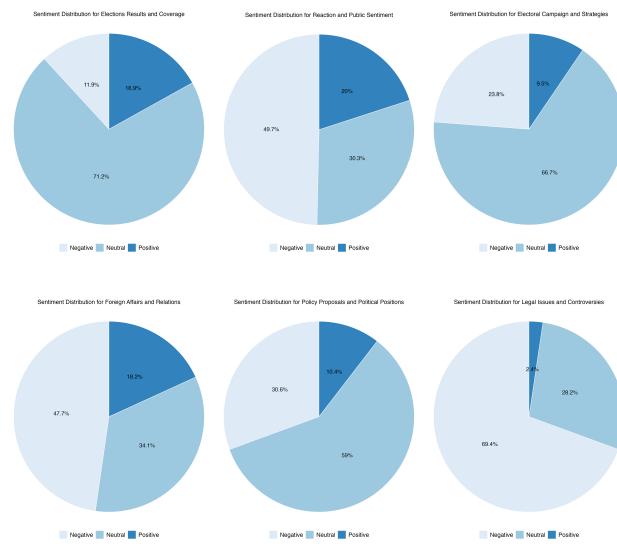
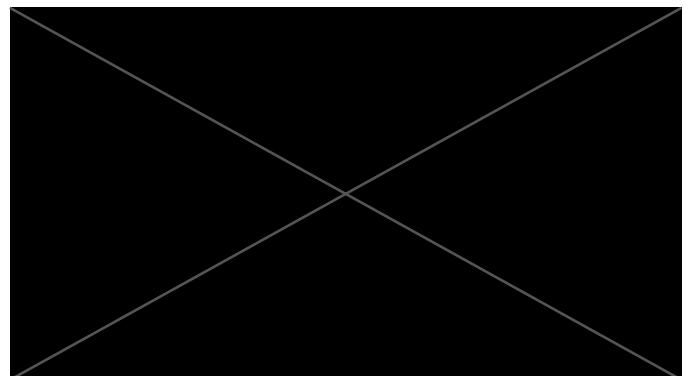


Figure 3: Pie Charts showing the distribution of sentiment annotations within each of the 6 thematic categories



References

- AllSides. 2024. Unbiased News: AllSides Balanced News & Media Bias Ratings. Retrieved from <https://www.allsides.com/unbiased-balanced-news>.
- Prokop, A. 2023. Why Iowa is So Important for 2024 Presidential Election. *The Week*. Retrieved from <https://theweek.com/news/world-news/us/962059/why-iowa-is-so-important-for-2024-presidential-election>.

Reuters. 2024. Trump nominates RFK Jr. to top U.S. health job. Retrieved from <https://www.reuters.com/world/us/trump-nominates-rfk-jr-top-us-health-job-2024-11-15/>.

Weissert, W.; and Gelston, D. 2024. Trump, Musk and new cabinet nominees celebrate at UFC. *BBC News*. Retrieved from <https://www.bbc.com/news/articles/ckgd1p8w4xlo>.