COMP 370 Final Project

Kamala Harris' Coverage In North American Media

Charlotte Marchal, Ilana Martin, Sarah Matmati

McGill School of Computer Science
charlotte.marchal@mail.mcgill.ca, ilana.martinmorales@mail.mcgill.ca,
sarah.matmati@mail.mcgill.ca

Introduction

On August 5, 2024, Vice President Kamala Harris replaced Joe Biden as the Democratic Party's nominee for the 2024 U.S. presidential election. Her 107-day campaign attempted to distinguish herself from Biden, assert her qualification for president over Donald Trump's, and align herself with the American populace. As the incumbent Vice President and the Democratic candidate in a historical presidential election, Kamala Harris figured heavily in North American political media coverage leading up to Election Day on November 5, and the weeks afterward that consolidated Trump's victory.

Our report examines both the spread of focal topics and the positive-negative skew, in line with our client's expressed desires, of 500 articles that mention Kamala Harris from English-language, North American online news outlets over a one month period: two weeks before the election, and two weeks after. This exploration of Harris's media coverage may also be of interest to those curious about the results of the 2024 election and the treatment of women in political media coverage, and those curious about the narrative factor of political campaigns.

Based on our collection and analysis of these articles, our key findings about Kamala Harris's coverage in the media specified above are the following:

- The majority of recent coverage of Kamala Harris is focused on neutrally updating on her role as the Democratic candidate in the 2024 U.S. presidential election, rather than on her current role as the Vice President.

- Donald Trump is a looming presence in Harris's coverage, both as the opposition to her campaign, in his position as the Republican candidate, and as a malicious commentator of Harris, through his frequent disparaging remarks against her.

- Endorsements of Harris, by figures in politics and entertainment, are a positive counterweight to the negativity in her coverage introduced by Trump's remarks.

- Mentions of Harris's character outweighed discussions of her policies, reflecting the media's focus on personal narratives.
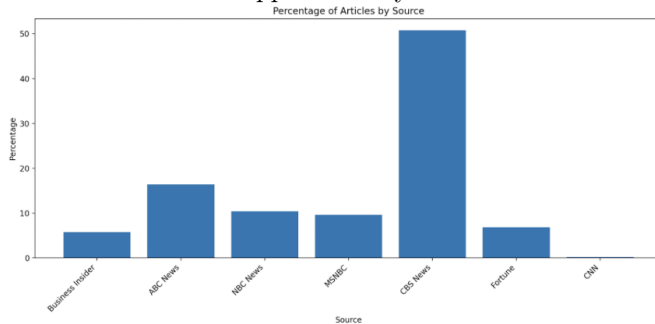
Data

In order to understand how Kamala Harris is being covered in the media, we built a dataset containing relevant articles. The dataset is composed of 500 articles about Kamala Harris obtained through Newsapi.org and includes 5 columns: Title, Description, URL, Label, Sentiment, defined as follows:

(1) Title: The title of the news article.

(2) Description: A snippet from the article as defined by Newsapi.org.

(3) URL: The URL to the news article, which also helped identify the article's source.

(4) Annotation: A human-annotated inferred feature indicating the topic of the article, selected from a typology guide we designed.

(5) Sentiment: A human-annotated inferred feature indicating the article's overall sentiment toward Harris.

To ensure the articles aligned with the data science question we aimed to address, we applied several criteria. First, to focus on articles specifically relevant to Kamala Harris rather than those that only briefly mentioned her, we filtered for instances where "Kamala Harris" appeared in either the title or the description. This was achieved using the API's "searchIn" parameter, where we were able to specify that the words "Kamala Harris", "Kamala" or "Harris" must appear in the title or the description of the article. Second, we needed all articles to be written in English and obtained from varied

reputable North American sources. We attempted to selected the best sources available but due to the limited availability of sources on Newsapi.org, we were restricted to the following sources: Associated Press, Bloomberg, Business Insider, CBS News, Fortune, Fox News, Google News, MSNBC, NBC News, The Wall Street Journal and The Washington Post. This filtering was accomplished using the "sources" and "language" parameters in the API. The following image demonstrates the distribution of articles' sources, with most news outlets having similar representations,except for CBS News, which accounts for approximately 50% of the articles.

It is important to note that the API's 30-day lookback window, and combined with the timing of the data collection, shortly after the United State Presidential Elections, resulted in a dataset biased toward election-related content.

Methods

Initial Attempts

When we initially began data collection, we attempted to research two lesser-known political figures. However, this proved challenging due to the API's limited lookback window and restricted availability of news sources, making it difficult to gather enough articles for meaningful analysis. To address this, we reflected upon current events that would likely generate substantial media coverage. As a result, we shifted our focus to a prominent and relevant figure: Kamala Harris.

Data Collection Methods

In our data collection, as outlined in the Data section, we focused on articles explicitly mentioning "Kamala Harris" in the title or description. This filtering was implemented to ensure relevance and reduce noise, as articles with only tangential mentions of Harris were less likely to provide meaningful insight into the media's portrayal of her. While this decision potentially excluded broader contextual discussions, it allowed for more precise and focused analysis.

Additionally, the dataset is inherently biased toward the 2024 Presidential Elections due to the timing of data collection and the 30-day lookback constraint of Newsapi.org. We opted to retain this bias, as election-related coverage likely represents the most current and impactful perception of Kamala Harris. This decision was made to prioritize timeliness and relevancy over an artificially balanced representation of media coverage across diverse topics.

Data Annotation and Typology

The data annotation involved two human-inferred features, Label and Sentiment, both detailed in the "Data" section.

We began with an open coding of 200 articles to determine the Label of each, which indicates the overall topic of the article, based on title and description. Initially, we had one team member annotate the first 200 articles, this resulted in a set of 10 categories. Then, as a group, we reviewed these categories and identified several areas for improvement: some categories were too similar and could be merged, others were overly specific and needed to be broadened, and we needed clearer labels for articles not directly focused on Harris. Through this process, we refined the categories and reduced them to a final set of 7, as follows: election, campaign, policies, endorsements, mentions of Kamala Harris by other political figures, Kamala Harris outside of her campaign and general facts about presidential elections. We then created a document with specific definitions and inclusion and exclusion examples for each type. Finally, we ensured the typology was comprehensive and well-defined by completing the labeling of the 300 remaining articles, among the 3 group members. We also went through all the annotations as a group to confirm the quality of the annotations and minimize human error.

As mentioned, we also annotated each of the 500 articles as "positive," "negative," or "neutral" based on their tone and content. During this annotation process, we carefully examined the language and diction employed in each article, noting how word choice and connotations contributed to the overall tone. Positive articles often used terms emphasizing competence and charisma, while negative ones frequently employed critical or skeptical language and were mostly from the perspective of other candidates. Neutral articles tended to focus on factual reporting with minimal evaluative language.

It is important to note that the labelling is limited by its reliance on snippet-based summaries from NewsAPI, which may not capture the nuance of full articles.

Tf-idf Analysis

To obtain the Tf-idf results from our CSV file, we began by writing a script to extract the elements for each category, as we aimed to identify the top ten words per category. This process resulted in a separate CSV file for each category. Since we split up the annotation, it was crucial to ensure consistency in the inputs. Therefore, the extraction process included data cleaning, such as handling extra spaces. To minimize errors and save time, we automated this cleanup within our script rather than performing it manually, which could have led to oversights. We considered removing the obvious keywords "Harris" and "Trump," which were likely to appear at the top of each category. However, we decided to keep them, as it was interesting to observe how these keywords ranked across different categories and to examine any variations in their prominence.

We developed a script to calculate the Tf-idf scores for the top ten words in each category, following the method discussed in class. These scores were ranked from highest to lowest and compiled into a single file organized by category. To clearly visualize the results, we used Python's Matplotlib library to create bar charts for each category. This visualization highlights the significance of each word within its category relative to others. Instead of converting the scores to percentages, we retained the raw Tf-idf values, as this approach provided clearer and more straightforward visualizations.

Results

Final Typology Definitions

As outlined in Methods, the final set of categories in our typology derived from the open-coding is as follows:

- **Election:** Articles providing predictions or updates about the 2024 U.S. presidential election, including but not limited to polls and vote counts for specific regions and demographics. An example of an article in this category is 'Harris concedes election, saying she promised Trump a peaceful transfer of power'; an example of an article *not* in this category is 'Trump and Harris make final appeals in Georgia as polls show tight race,' since it pertains to the candidates' campaigns.

- **Campaign:** Articles relating to the rallies, interviews, and other appearances and statements Kamala Harris made as part of her campaign for the 2024 presidential election. An example of an article in this category is 'Harris campaigns in Michigan, works to shore up support with Black men'. An article that is *not* in this category is 'Harris calls Trump

to congratulate him on 2024 election win, aide says', since at this point, her campaign has ended.

- **Endorsement:** Articles about explicit support of Kamala Harris for president by influential figures such as musicians, politicians, journalists. This includes opinion pieces like 'I'm a lifelong Wisconsin conservative. Here's why I just endorsed Kamala Harris,' as well as articles about Harris campaigning *with* a figure who endorsed her: 'Harris and Obama campaign together for first time'.

- **Mentions of Kamala Harris by other political figures:** Articles about statements made by other politicians about Kamala Harris and her affiliated party, such as about her running mate Tim Walz. An article in this category is 'The irony of Trump accusing Harris of having 'the plans of a simpleton', while an article that is *not* in this category is 'Harris to campaign with Obama, Springsteen; Trump heads to the West', since although Obama is a political figure, the article is moreso related to an event on her campaign.

- **Kamala Harris outside of her campaign:** Articles which relate to Kamala Harris's current role as the Vice President as well as other characteristics of hers that do not directly relate to her 2024 campaign. An example of an article that is *not* in this category is 'What we know about Trump, Harris medical records after VP releases doctor's report' since although it is about personal information about Harris, it is moreso related to her campaign against Trump for the election.

- **Policies:** Articles about Kamala Harris's policies she would enact as president, including her proposed cabinet. This includes articles about voter concerns regarding specific topics which Harris would address through policies as president. For example, '"Three Meals": Arizona voters weigh in on immigration' is in the Policies category.

- **General facts about presidential elections:** Articles relating to facts about and analyses of U.S. presidential elections in general, not specifically the 2024 election. For example, articles about the importance of a certain demographic's vote are included in this category, since that information is consistent for multiple elections, such as 'Examining the impact of the Latino vote in the 2024 election'.

Annotation Results

The annotation of the articles resulted in the following distribution of labels from the typology:

Distribution of Annotations
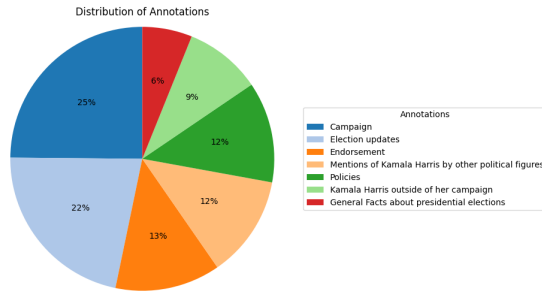
The dominating categories are 'Campaign', with 25% of the articles, and 'Election updates', with 22%.

The least populated categories are 'General Facts about Presidential elections', with 6% of the articles, 'Kamala Harris outside of her campaign', with 9% of the articles, and 'Policies', with 12% of the results.

Tf-idf Scores

The distribution of Tf-idf scores is relatively uniform for categories like "General Facts About Presidential Elections" and "Kamala Harris outside of her campaign", where the scores decline gradually. In contrast, the distributions are more sharply skewed in categories like "Campaign" and "Election Updates", where the top-scoring words dominate significantly.

The categories with the highest raw Tf-idf scores are "Campaign", "Election Updates", and "Policies", indicating a stronger relevance of specific terms in these areas. Conversely, categories like "General Facts About Presidential Elections" and "Endorsement" show lower maximum scores, suggesting a more balanced importance across terms.
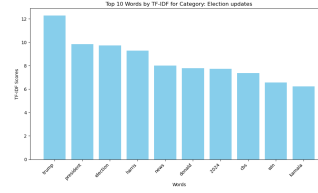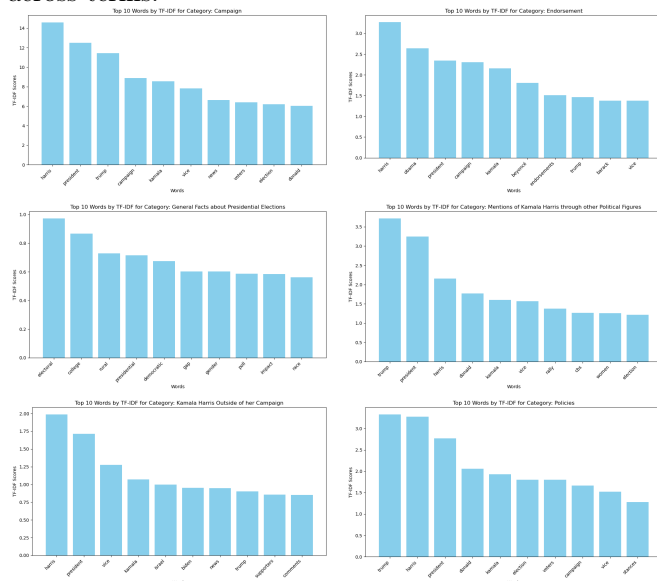
Sentiment Analysis

The sentiment analysis resulted in 64% of the articles being labeled as neutral; 20% as positive and 16% and negative. The following graphs show the proportions for each category, as well as overall. Most of the categories had similar proportions of positive, negative, and neutral articles, with a majority of neutral articles, and then a fairly even split between positive and negative articles. We note three categories that are exceptions to this: 'Endorsement', which had a majority of positive articles; 'Mentions of Kamala Harris by other political figures', which had a majority negative; and 'General facts about presidential elections', which were about 91% neutral and 9% negative:
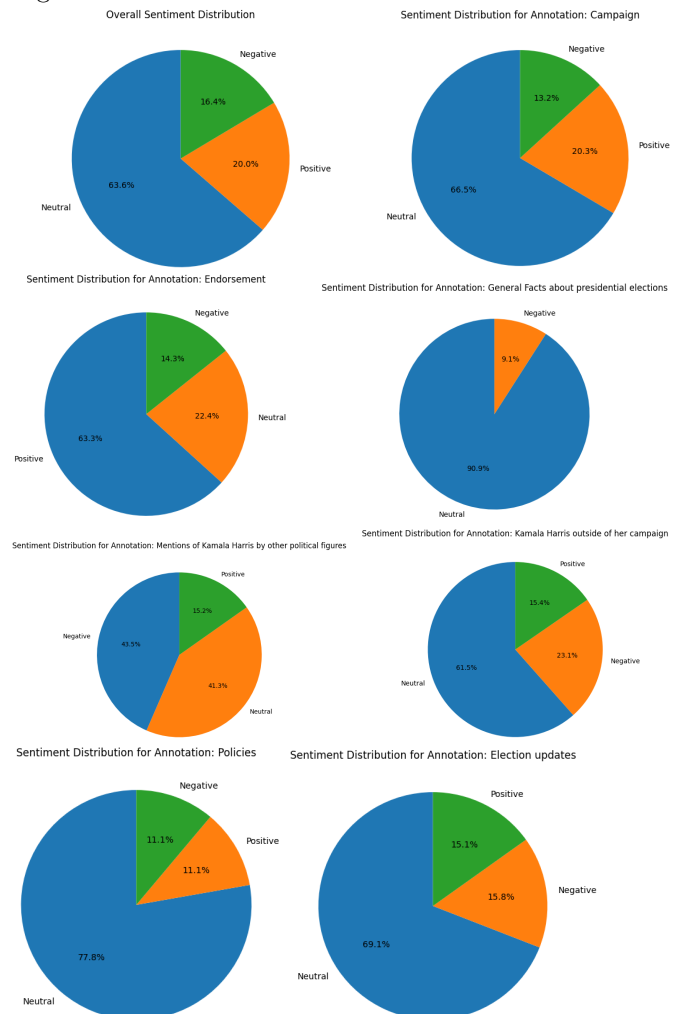
Discussion

Composition of Categories

The majority of the English-language, North American online news coverage of Kamala Harris is focused on the 2024 U.S. presidential election, with all but one category in our typology, 'Kamala Harris outside of her campaign,' encompassing articles pertaining to it. Moreover, this category, which encompasses Harris's role as the current VP, is sparsely populated (containing 3% of the 500 articles), indicating the precedence her role as a candidate took in the news around the time of the election.

Furthermore, relatively few articles focused on her proposed policies (2%), while mentions of her character, in the form of updates on campaign activities, such as rallies and speeches, were more prominent: the 'Campaign' category is the largest, with 24% of articles. This suggests that the media tends to gravitate toward personal narratives over important policy discussions, reflecting that the personability of a candidate attracts greater public and media attention than the governing choices they represent.

Tf-idf Analysis

The tf-idf analysis revealed many interesting insights about Kamala Harris's coverage in North American media regarding endorsements by political and entertainment figures, the relevance of international conflicts to American politics, what factors are influential to the outcomes of presidential elections, and the comparative nature of representative democracy.

Firstly, we note the high scores for 'Obama' (2.6414) and 'Beyoncé' (1.8039) in the coverage of Kamala Harris within the 'Endorsement' category. The prominence of this terms indicates the weight the collected news sources put on these figures endorsing her. As the first Black President, the support of Obama for Harris, whose Black identity was continuously questioned by opposition, may have been instrumental in affirming her identity to skeptical voters. Beyoncé's endorsement of Harris may have been pertinent for similar reasons, as well as her status as a veritable popstar: we see that the news places significance on the political opinions of entertainment figures rather than other political figures, suggesting that in the current political climate, so do voters. Notably absent from the top-scoring word list in this category is 'Biden.' This absence may indicate a deliberate effort, either by Harris, Biden, or the media, to distance her from him following his exit from the candidacy, despite their close working relationship.

Another interesting observation is within the "Kamala Harris outside of her Campaign" category: the prominence of the word "Israel," which ranks among the top ten words with a raw tf-idf score of 0.9971. This suggests that "Israel" is a significant term in the category-specific data, indicating its importance and distinctiveness compared to other key terms. While expected words like "president," "vice," "Kamala," and "Biden" reflect central figures and roles directly tied to the overarching theme of our research, the prominence of "Israel" stands out. This may point to a specific focus on international relations involving Kamala Harris, particularly in her capacity as Vice President, rather than in the context of her campaign activities. Overall, this highlights the profound influence of real-world events on political discourse, illustrating how relevant and impactful global-political issues can be on the perception and roles of presidential candidates.

The "General Facts About Presidential Elections" category is notable as the only one which does not include 'Kamala' in its top words, nor any other political figure. The heavyweight topics include the Electoral College, which is the process that determines the outcome of the election as opposed to the popular vote; how voters in rural areas affect the election outcome; and how gender plays a role in elections. This makes sense because the "General Facts About Presidential Elections" category focuses on structural and procedural aspects of elections rather than specific individuals or their roles. This category reflects discussions about the mechanics of the election process and broader societal dynamics. Another interesting keyword in the "General Facts About Presidential Elections" category is "race," which introduces nuance. This term can include articles about the roles different racial groups play in the election, as well as the candidates' campaigns being a 'race'.

Perhaps the starkest result from the tf-idf analysis is the significant presence of the term "Trump" across all categories. In some cases, this is expected, for example, in the 'Election Updates' category, where the election results naturally highlight Trump's prominence. However, it is more surprising in categories like "Kamala Harris Outside of Her Campaign," where Harris would be expected to be the primary political figure discussed. This anomaly likely reflects the time frame during which the articles were written, aligning with the election cycle, where Donald Trump, as a central political figure and opponent, was frequently mentioned with Kamala Harris. This overlap in coverage underscores the polarized nature of election discourse and the media's tendency to frame narratives through comparisons between candidates.

Sentiment Analysis

Endorsements of Harris are overwhelmingly positive (63.3%), a result that aligns with the purpose of endorsements, which are typically expressions of agreement, support, and approval of a candidate's vision and policies.

On the other hand, mentions of Harris by other political figures are notably more negative (43.5%). This is likely driven by criticism from opponents, particularly Donald Trump, who played a central role in shaping the discourse around her during the election cycle. The negative tone in this category underscores the adversarial nature of political campaigns, where opponents aim to challenge each other's qualifications and track records.

In contrast, neutral sentiment dominates categories such as election updates (69.1%), discussions of policies (77.8%), and general facts about presidential elections (90.9%). This reflects the largely informational tone of articles in these areas, where the focus is on delivering factual, objective content rather than opinion or commentary. These high levels of neutrality are indicative of the media's role in providing updates and context for the public, especially during a politically charged time like an election.

Conclusion

The majority of Kamala Harris's North American online news coverage from the month surrounding Election Day is neutral: primarily, articles that mentioned her were unbiased updates on her campaign for President, with a smaller subset outlining her policies she would implement. Non-neutral articles were mostly positive; however, this is thanks to those that documented her endorsements, rather than being positive op-eds. Similarly, most of the negative coverage of Harris were articles about Trump's derisive remarks towards her, rather than journalistic critiques of her candidacy or role as Vice President. These results speak to the desire for media coverage to be unbiased in tone, which can contribute to an air of credibility, and enable the readers to draw conclusions themselves.

It is important to note the existence of another angle for media sentiment regarding Harris that we consciously did *not* look through in our analysis: the fact that the choice to cover one topic over another is not a politically neutral decision for news outlets. We abstained from this angle because what topics are considered positive and negative differs based on political affiliation. For example, a left-leaning review of Harris's coverage would likely view a large number of articles about Harris's stance on abortion as positive, while a right-leaning would view it as negative. An analysis considering this extra vantage point would require more conversation with the stakeholders.

For a broadened sense of the media sentiment for Kamala Harris, we would also suggest an expanded foray into other media sources, such as social media, where positive and negative evaluation of Harris may be more apparent due to the candid nature of the posts, as opposed to traditional news outlets, which prefer to appear unbiased.

Group Member Contributions

Charlotte Marchal: Performed open coding of 300 articles, developed typology and guide, performed sentiment analysis of 1/3 of the articles, wrote sections "Introduction", "Final Typology Definitions", "Discussion" and "Conclusion" of the report.

Ilana Martin: Wrote script to extract articles, performed open coding of 100 articles, performed sentiment analysis of 1/3 of the articles, wrote sections "Data", "Methods, "Sentiment Analysis" and "Discussion" of the report.

Sarah Matmati: Performed open coding of 100 articles, performed sentiment analysis of 1/3 of the articles, wrote script to calculate td-idf scores, wrote sections "Methods", "Tf-idf Scores" and "Discussion" of the report.

References

Canadian Broadcasting Corporation. (2024, October 1). Netanyahu warns Israel may strike Iran over nuclear threats. CBC News. https://www.cbc.ca/news/world/netanyahu-israel-strike-iran-nuclear-1.7386269

CBS News. (2024, October 1). Kamala Harris warns Iran not to respond to Israel counterstrike. https://www.cbsnews.com/news/kamala-harris-warns-iran-not-to-respond-to-israel-counterstrike/

National Archives and Records Administration. (2024, November 4). What is the Electoral College?. National Archives and Records Administration. https://www.archives.gov/electoral-college/about

NPR Washington Desk. (2024, August 6). It's official: Kamala Harris becomes Democrats' 2024 presidential nominee. NPR. https://www.npr.org/2024/08/01/nx-s1-5060456/harris-democratic-nominee-roll-call-delegates

The White House. (2024, October 1). Remarks by Vice President Harris on the Iranian ballistic missile attack against Israel. https://www.whitehouse.gov/briefing-room/speeches-remarks/2024/10/01/remarks-by-vice-president-harris-on-the-iranian-ballistic-missile-attack-against-israel/