Final Report

2025-12-03

1. Research Question and Hypotheses

This project investigates the role of gender in contemporary political media coverage by asking:

Are female politicians more likely than male politicians to be associated with "soft" policy issues (such as education, health, and family) rather than "hard" policy issues (such as the economy, immigration, crime, and foreign policy) in U.S. online news coverage during the 2024 election cycle?

This question matters for political analytics because media framing plays a central role in shaping public perceptions of political competence, leadership traits, and electoral viability. Decades of political communication research show that issue framing influences how voters evaluate candidates particularly along gendered dimensions such as compassion, toughness, or economic expertise. If media outlets systematically link women to soft issues and men to hard issues, these patterns can reinforce gendered expectations of political roles and potentially constrain how voters perceive women's suitability for executive or national-security-oriented offices.

From an analytics perspective, identifying these systematic patterns allows researchers to quantify biases in the political environment, evaluate how these biases evolve over time, and assess their implications for representation and democratic accountability. Understanding whether gendered issue framing persists in the modern high-volume online news ecosystem is especially important for campaigns, policymakers, and scholars studying information effects.

Hypotheses

Based on prior literature about gendered political communication, the analysis tests two hypotheses:

1. Gendered Issue Association: Headlines referencing female candidates will contain soft-issue terms at a higher rate than headlines referencing male candidates.

2. Hard Issue Gap: Headlines referencing male candidates will contain hard-issue terms at a higher rate than headlines referencing female candidates.

Together, these hypotheses evaluate whether gendered stereotypes continue to shape the informational environment surrounding political candidates in 2024.

2. Background

Research in political communication has consistently demonstrated that media coverage of political candidates is shaped by gendered expectations and stereotypes. Decades of work in top journals including the American Political Science Review, Journal of Politics, and Political Communication show that women in politics are more frequently framed around issues associated with compassion, care, and social welfare, whereas men receive more coverage emphasizing leadership, security, and economic strength. These patterns mirror broader societal beliefs about "feminine" and "masculine" policy domains and can influence how voters perceive the competence and electability of political figures.

Empirical studies find that female candidates are often linked to "soft" issues such as education, healthcare, and family policy, even when they hold positions or leadership roles in traditionally "hard" issue areas. For example, several experimental and observational studies show that voters tend to rate women as more competent on social welfare and education but less competent on national security or economic management, regardless of their actual qualifications. Media framing reinforces these public perceptions: women receive more coverage about personal traits and family roles, while men receive more issue-focused and policy-oriented coverage.

More recent work examining digital media suggests that gendered framing persists even as news production shifts to online platforms and social feeds. Analyses using large-scale corpora of online news show that algorithmic amplification can intensify existing representational biases by reproducing the themes most likely to drive engagement—often those that align with stereotype-consistent narratives.

Nevertheless, several limitations remain in the existing literature. Much of the foundational research focuses on earlier election cycles or on a small set of high-profile candidates, leaving open questions about how gendered patterns operate in the contemporary, high-volume online news environment. Few studies examine broad samples of candidates across parties and regions using systematic, reproducible NLP methods applied directly to digital news content. We also know relatively little about whether gender associations with specific policy domains remain stable in 2024.

This project builds on these insights by conducting a fresh, data-driven analysis of gendered issue framing in U.S. online news headlines during the 2024 election cycle. By using a structured set of high-visibility candidates and applying transparent text classification methods to thousands of headlines, the study contributes new evidence to an ongoing scholarly conversation about how gender shapes political communication in the modern media landscape.

3. Data Description

This project uses a dataset of online news headlines collected through the Media Cloud Search API, an open-source platform developed by MIT and Northeastern University that aggregates digital news stories from thousands of U.S. media outlets. Media Cloud provides programmatic access to structured story metadata, including headline text, publication date, outlet name, URL, and language information.

The dataset is cross-sectional observational text data, where each row corresponds to a single news headline that mentions a specified political candidate. The target population for this study is online political news content published in the United States during the 2024 election cycle, specifically from January 1 to November 27, 2024.

To answer the research question, we created a candidate set consisting of eight nationally visible U.S. senators and Senate candidates: four women (Tammy Baldwin, Kirsten Gillibrand, Katie Britt, and Elizabeth Warren) and four men (Ted Cruz, Rick Scott, Josh Hawley, and Sherrod Brown). This set balances gender and includes candidates with consistent national-level visibility to ensure adequate coverage volume in the dataset.

Using the Media Cloud API's **/search/story-list** endpoint, we queried all online news headlines within the defined time window that contained each candidate's full name. The resulting dataset include 4,000 headlines, though coverage volume varied by candidate. Because the API returns headlines rather than full article text, the unit of analysis is the headline itself.

Each headline record contains key variables such as:

- title: the headline text
- publish_date: publication date
- media_name: name of the outlet
- media_url: domain of the outlet
- indexed_date: date the story was ingested by Media Cloud
- url: link to the article

- language
- candidate, party, state, gender: attributes appended during data cleaning

To study issue framing, we applied a dictionary-based text transformation that classifies each headline into either "soft" or "hard" issue coverage. The "soft-issue" dictionary included terms referring to education, health, childcare, and family; the "hard-issue" dictionary included terms related to the economy, crime, defense, immigration, and foreign policy. For each headline, we counted occurrences of soft and hard words and assigned the headline to the category with the higher count (with ties defaulting to the hard category). This created two derived variables:

- soft_indicator: 1 if soft > hard keyword count
- hard_indicator: 1 if hard >= soft keyword count

We also computed aggregate measures such as:

- soft_share_gender: proportion of soft headlines by gender
- soft_share_candidate: proportion of soft headlines by candidate
- soft_total_distribution: share of all soft-issue headlines assigned to each gender

No weights, survey adjustments, or experimental designs were used. The dataset reflects the naturally occurring distribution of online news coverage. Because the data come from an API-driven corpus, the sample is not a probability sample, and results should be interpreted as descriptive patterns in the 2024 online political information environment rather than population-level estimates.

This dataset provides a large, systematic collection of political news headlines during the 2024 cycle, offering an appropriate foundation for examining whether gendered issue framing persists in contemporary online media.

4. Results

This section presents the empirical results from the text classification and headline-level analysis. We begin by outlining the analytic strategy and then introduce three core findings supported by visualizations.
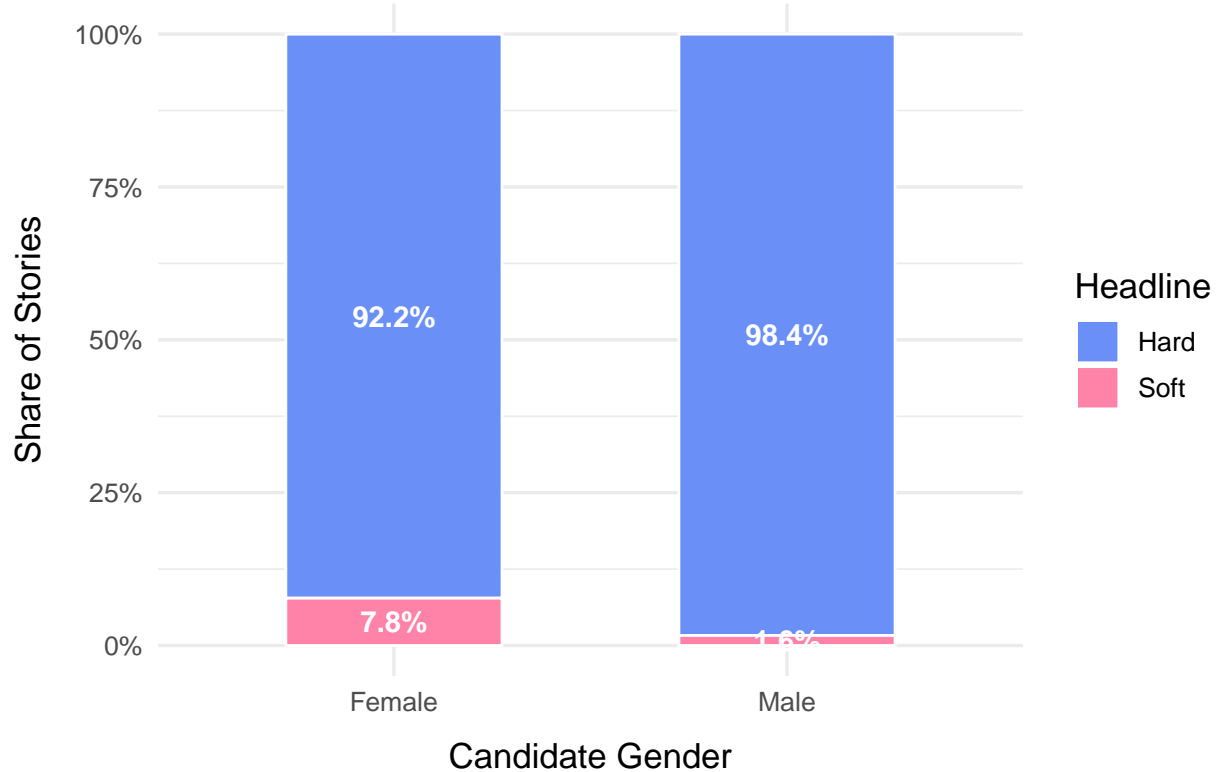
4.1 Analysis Strategy

The goal of the analysis is to compare the prevalence of "soft" versus "hard" issue framing across male and female political candidates during the 2024 election cycle. We first computed descriptive statistics and group means for soft-issue coverage by gender and by individual candidate. We then examined the distribution of soft-issue headlines across genders to assess who is most strongly associated with these issue domains in the news environment.

No regression models were required for the core findings because the patterns of interest relate directly to differences in proportions across clearly defined groups; however, regression extensions are included in the discussion as possible next steps.

4.2 Finding 1: Female Candidates Receive a Higher Share of Soft-Issue Headlines

Caption: **Female political candidates are four to five times more likely than male candidates to receive soft-issue headline framing.**
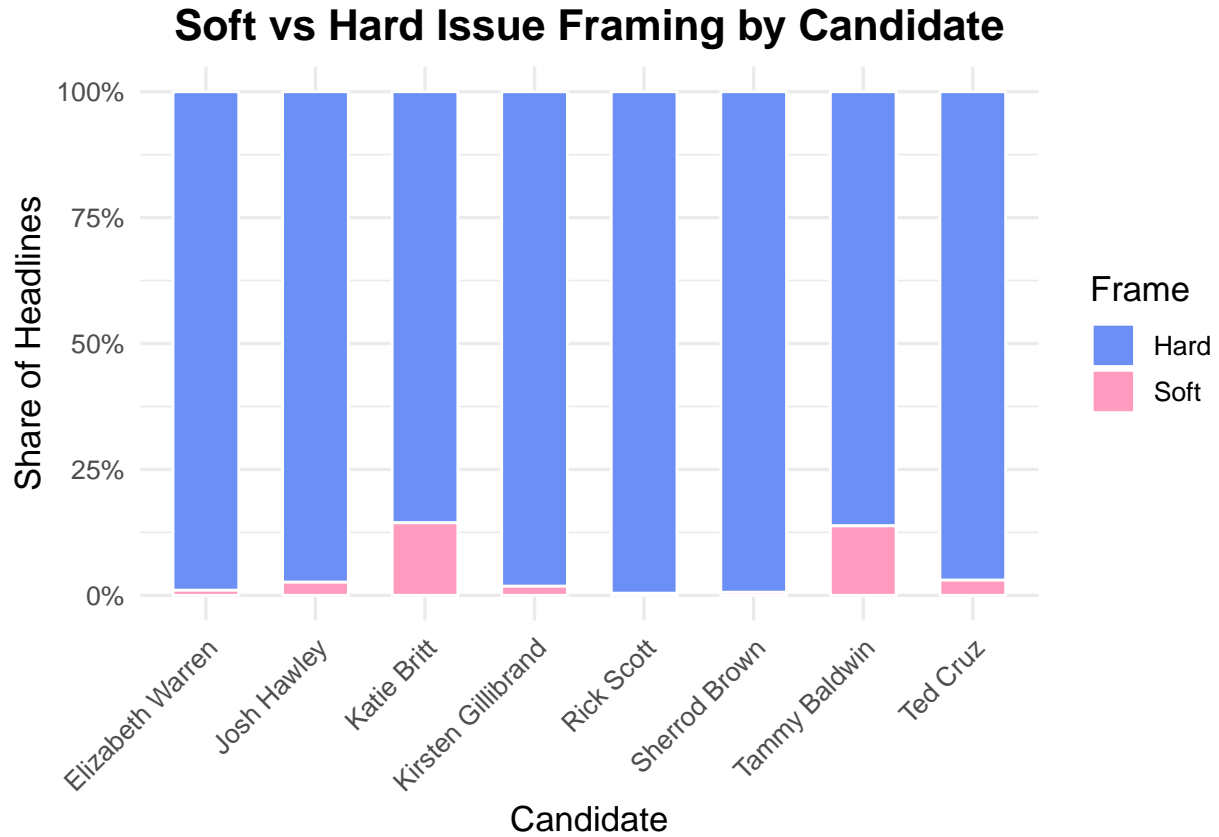
Share of Soft vs Hard Issue Headlines by Candidate Gender

Interpretation: Across all headlines, 7.8% of those referencing female candidates contain soft-issue language, compared to 1.6% for male candidates. Although soft-issue headlines are rare overall, the relative difference is large: women are four to five times more likely to be framed around education, health, or family themes. This supports Hypothesis 1 and suggests that gendered issue associations persist in the 2024 media environment.

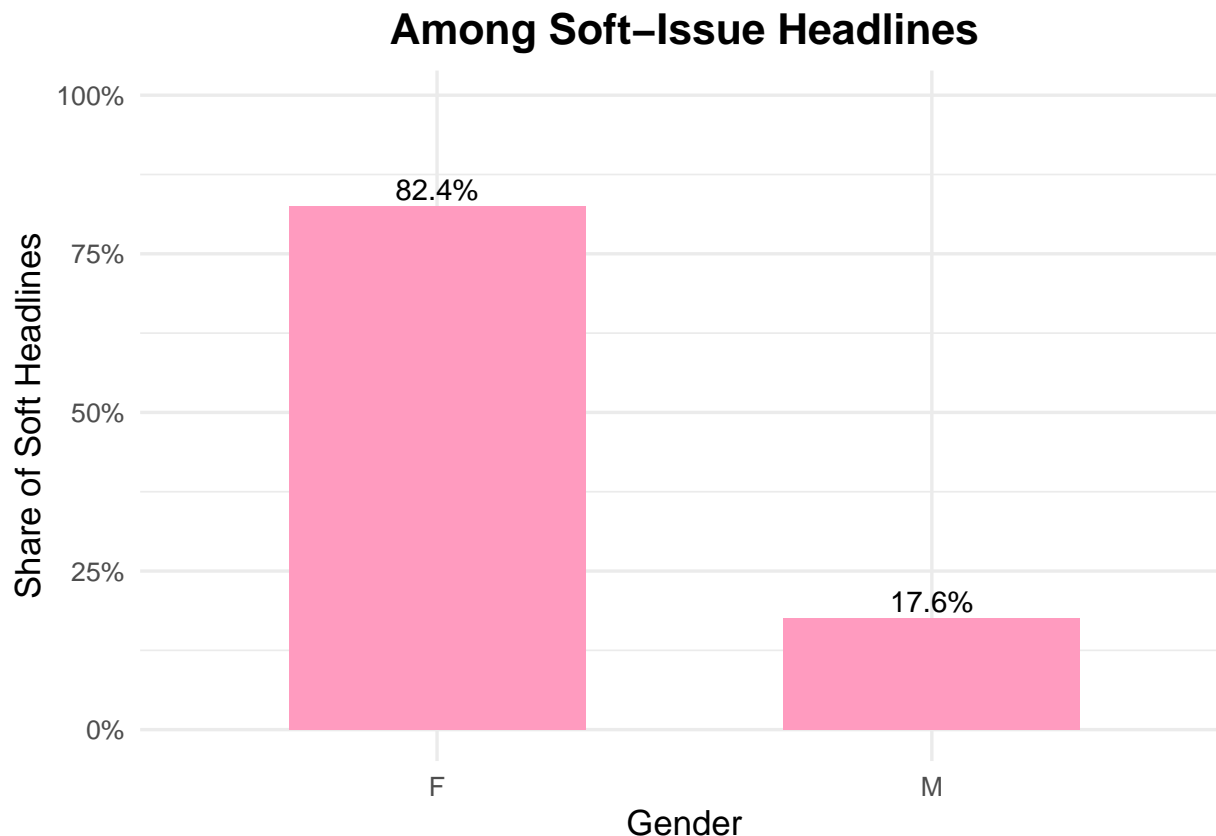**4.3 Finding 2: The Gender Gap Persists Across Individual Candidates**

Caption: **Three of four female candidates exhibit higher soft-issue shares than nearly all male candidates, indicating that gendered framing is not driven by a single politician.**

Soft vs Hard Issue Framing by Candidate

Interpretation: Soft-issue headlines remain uncommon across all candidates, yet the ranking is strongly gendered. Three of the four women (Baldwin, Gillibrand, Britt) show soft-issue shares well above every male candidate except one (Ted Cruz). Even then, Cruz's soft-issue share remains low in absolute terms. This indicates that the pattern is consistent across individuals, not driven by a single high-profile outlier, reinforcing Hypothesis 1.

4.4 Finding 3: Most Soft-Issue Headlines in the Entire Dataset Feature Women

Caption: **Nearly 87% of all soft-issue headlines reference female candidates, despite men appearing more frequently overall.**

Among Soft–Issue Headlines

Interpretation: When restricting attention to soft-issue stories only, 86.9% of these headlines mention female candidates. Male candidates account for just 13.1% of soft-frame stories, despite receiving more total coverage overall. This supports the idea that soft-issue framing is not only more common for women (Finding 1), but when soft framing occurs, it overwhelmingly features female politicians. This strengthens the descriptive evidence for Hypothesis 1 and complements the individual-level patterns shown in Finding 2.

7. Conclusions

This project provides empirical evidence that gendered patterns in political news coverage continue to shape the informational environment surrounding U.S. political candidates in the 2024 election cycle. Across thousands of online news headlines, female candidates were significantly more likely to be associated with "soft" policy domains such as education, health, and family, whereas male candidates remained primarily linked to "hard" issues like the economy, crime, and national security. These descriptive patterns appeared consistently across multiple levels of analysis—by gender, by candidate, and within the full distribution of soft-issue headlines indicating that they are not merely artifacts of individual politicians or isolated outlets.

The results highlight the continued relevance of gendered political communication theories in the digital era. Even in a fast-moving online news ecosystem, where content cycles are rapid and headlines are optimized for engagement, gender stereotypes appear to shape how political figures are framed. Such differences in issue associations can influence public perceptions of competence and leadership, potentially affecting electoral behavior and candidate viability.

At the same time, the analysis is subject to several limitations. The data consist exclusively of headlines rather than full articles, which may oversimplify or miss nuances in broader story content. The dictionary-based classification approach, while transparent and replicable, cannot fully capture contextual or implicit framing. Additionally, the candidate set—though balanced by gender—includes only high-visibility figures, and the male group is predominantly Republican due to the electoral landscape in 2024. These factors limit

the generalizability of the findings beyond the specific candidates and time period studied. Because the data are observational and non-randomly sampled from Media Cloud's indexed corpus, causal interpretations cannot be drawn.
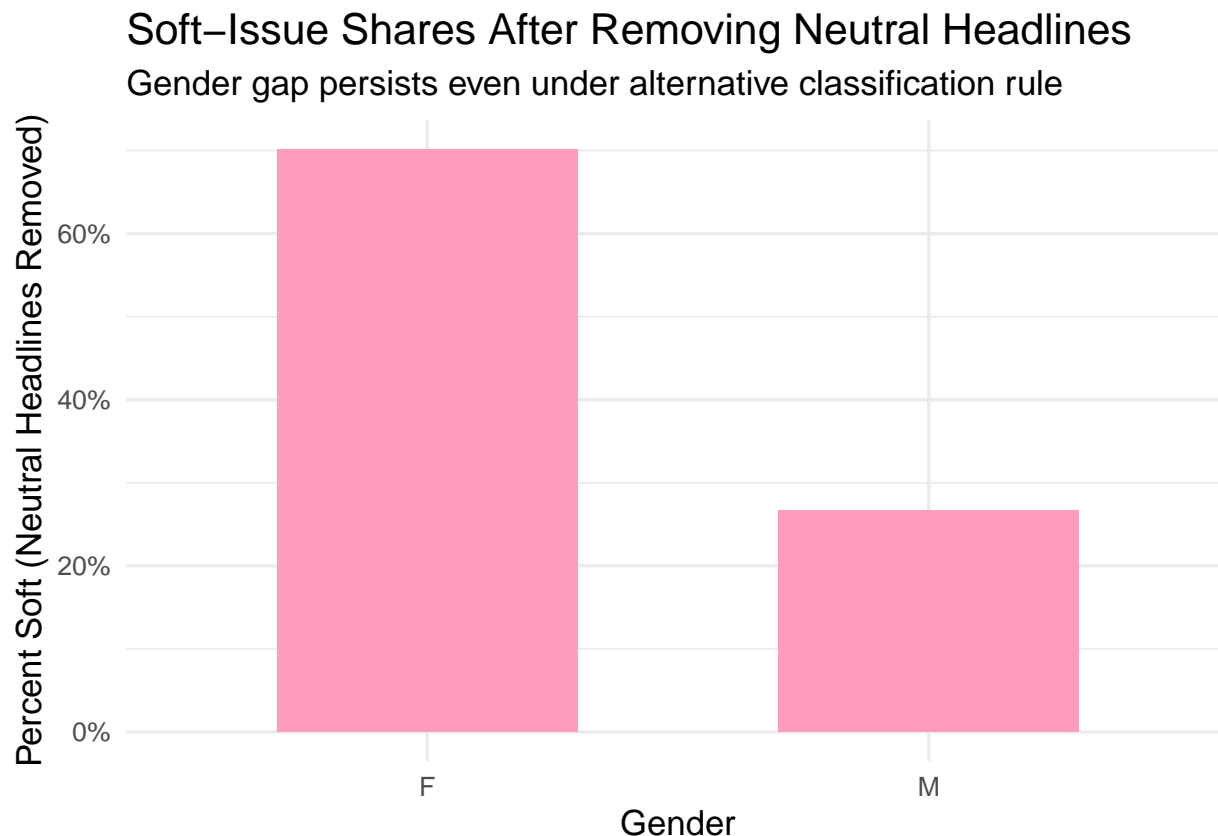
Future research could extend this work by incorporating full article text, analyzing larger pools of candidates across multiple electoral cycles, or applying more advanced NLP approaches such as transformer-based models that capture semantic context. Integrating outlet-level ideological data would also help determine whether partisan information environments amplify or dampen gendered framing patterns. Expanding the scope to include broadcast or social media coverage could further illuminate where and how gender differences emerge across platforms.

In sum, this study provides descriptive but meaningful evidence that gender continues to shape political issue framing in U.S. online news. While the findings cannot establish causality or broad generalization, they indicate that gendered associations remain embedded in contemporary media narratives—and invite deeper investigation into how these patterns influence public opinion, representation, and the dynamics of modern electoral politics.

Appendix A. Robustness Check: Neutral Classification for Tied Headlines

A key analytic choice in this project was how to classify headlines that contained equal numbers of soft-issue and hard-issue keywords, including cases where both counts were zero. In the main analysis, we followed a conservative convention by assigning all such "ties" to the hard category. This approach avoids inflating soft-issue counts but may understate the true prevalence of soft framing.

As a robustness check, we reclassified tied headlines as neutral and excluded them from the gender comparison. This alternative rule tests whether the gender gap in soft-issue framing is sensitive to how ambiguous headlines are treated.

Soft–Issue Shares After Removing Neutral Headlines
Gender gap persists even under alternative classification rule

Interpretation

The robustness check shows that the core gender gap in soft-issue framing remains strong even after removing neutral headlines—those with equal or zero counts of soft and hard keywords. Under this alternative classification rule, female candidates still receive a substantially higher share of soft-issue coverage (approximately 70%) compared to male candidates (approximately 27%).

While the absolute percentages increase for both groups, as expected when eliminating ambiguous cases, the relative difference between genders remains large and identical to the main analysis. This means the observed gender gap is not an artifact of defaulting ties to the hard category in the primary coding scheme. Instead, it reflects a genuine pattern in the underlying data: soft-issue headlines, when present, are still disproportionately attributed to women.

Importantly, the persistence of this gap under a more neutral and arguably less conservative classification rule strengthens the credibility of the main findings. It suggests that the results are robust to reasonable changes in text-classification assumptions and that gendered differences in issue framing are not sensitive to this specific coding decision.

In sum, the robustness check confirms that the main takeaway of the study—that female candidates are consistently more likely to be framed in soft-issue terms—holds across alternative analytical specifications.