Podcasting the News – A Topic, Sentiment, and Stance Analysis of U.S. Podcasts and Public News Media

SIADS 699 Capstone Project Final Report

Sean Fontaine, Chelsea Simpson

1 Introduction

The digital age has profoundly changed the way in which individuals consume news content. Formerly the exclusive domain of traditional publishers, the past two decades have seen alternative digital media occupy a rapidly expanding portion of the contemporary news media environment. While the rise of news on social media is well studied, podcasts as emerging sources of news information and commentary are less well understood. Podcasts have recently overtaken newspapers as a source of news amongst American consumers (*Newman, N.*, 2025); over 50% of Americans listen regularly, and 87% of listeners say they trust their podcast's news content (*Pew Research Center*, 2023).

This project investigates the central question:

In what ways does the news content presented in popular podcasts systematically differ from that produced by traditional public broadcasters?

In answering this question, we look at:

a) What topics are being talked about and how do topic distributions vary across mediums?
b) How do stances, sentiments, and framings of key topics compare across mediums?

Through this study, we aim to provide an evidence-based understanding of the role that podcasts play in the modern information landscape.

2 Related Work

The recent rise of podcasts as primary news sources (Newman, 2025; Pew Research Center, 2023) has left scholarly analyses of their news content underdeveloped. Existing research on political podcasts tends to focus on more granular podcast messaging dynamics rather than larger-scale analyses of podcast content. A leading study of this type by DeMets and Spiro (2025), *Podcasts in the periphery: Tracing guest trajectories in political podcasts*, uses network analysis (i.e. network centrality; Louvain methods for communities) to map how guests on podcasts influence messaging on political and top polarizing issues. Our study expands on current research by shifting the focus from messenger to the messaging itself by analyzing topics, sentiment, stances, and framing within the wider podcast-news information ecosystem.

Our methods are grounded in established techniques for analyzing news media content. Sentiment analysis is commonly used in media studies as a metric to inform analysis of framing and tone. We build on the sentiment analysis approach taken by Yu & Yang (2024), whose study uses sentiment as a metric to understand coverage of economic topics during the COVID-19 pandemic. To understand *what* is being discussed, we use BERTopic, a leading topic modeling technique that leverages embeddings to generate more context-aware and interpretable topics (Grootendorst, 2022). Furthermore, to understand *how* topics are being discussed, we draw on a novel Chain of Stance (CoS) framework (Ma et al., 2024), an LLM-based prompting method that has recently outperformed leading stance detection models on SemEval-2016, a benchmark stance detection dataset.

By greatly extending the scope of existing research and combining state-of-the-art media research techniques, our study offers a new and far more comprehensive view into the podcast-news information ecosystem.

3 Methodology

3.1 Data

We compiled a complete corpus of all content produced over the calendar year from July 1, 2024- July 1, 2025 by leading podcasts and public news outlets.

Podcasts

We select all podcasts in the Top 50 most-listened-to podcasts in the United States (*Edison Research*, 2025) during Q1 of 2025, excluding those affiliated with a traditional news outlet like *The Daily*, by The New York Times (see *Appendix A*). Transcripts are downloaded from *Podscribe* (transcription using Google Cloud's Speech-to-Text) with timestamps and speaker tags removed.

Public News Sources

To serve as a ground comparison for our podcasts, we select NPR and PBS, the two largest public broadcasters in the United States. Articles are scraped and html parsed from the archive of each site.

The assembled dataset contains 8,344 podcast transcripts from 40 publishers, and 17,435 news articles from PBS and NPR.

3.2 Topic Modeling

To gain an understanding of *what* is being talked about in our corpus, we employ topic modeling, an unsupervised machine learning method that extracts core topics within a source using clusters of words and phrases that tend to co-occur. We do this with BERTopic (Grootendorst, 2022), an embeddings-based approach that accounts for the context of words as well as raw frequencies to generate more sophisticated human-interpretable topics. Prior to modeling, corpus documents are split into chunks (of approx 300 words) to allow the model to focus on smaller, more cohesive units of information. Results of modeling are 402,415 chunks assigned to 3,232 unique topics based on semantic similarity.

3.3 Topic Labeling

Next, we label all found topics. Labels are assigned using The International Press Telecommunications Council's (IPTC, 2025) Media Topics, a standardized, hierarchical, and widely used media classification schema. This assignment is done by transforming IPTC topic labels and descriptions into embeddings using our BERTopic model's transformer, and matching by highest cosine similarity. All topics are then human reviewed for accuracy, and changed when needed.

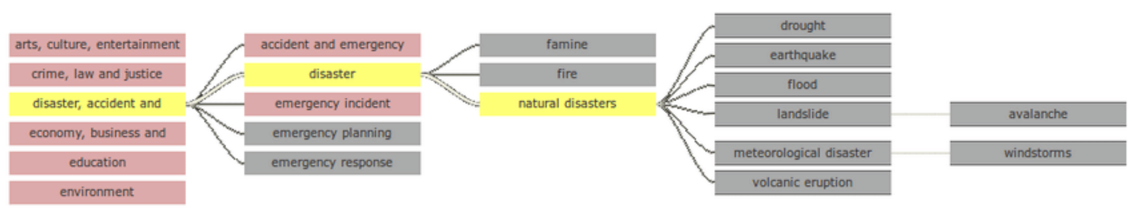

Figure 1: Excerpt of the IPTC Media Topics taxonomy (reproduced from Rudnik et al., 2019, Figure 3).

An alternative labeling schema is also applied to identify specific named entities (i.e. people, countries, and events). This is applied to chunks using simple string matching in a non-mutually exclusive manner.

The result of this is 1 hierarchically grounded IPTC media topic and up to 46 named entity topics for each of our 402,415 chunks.

3.4 Stance Detection

Stance detection is carried out using the Chain of Stance (CoS) prompting method (Ma et al., 2024). The method sequentially guides an LLM through six logical steps, gathering information and evidence about context, main idea, and tone before outputting a stance determination (see Appendix B for prompting details).

To implement this, the corpus is first filtered to 70 topics relevant to stance labeling (Appendix C). Inferences are then run on each document–topic pair in batches of 16 using an open-source LLM hosted on an AWS EC2 G5.xlarge instance. As Mistral models are open-source and have achieved top results in leading stance-detection research (Ma et al., 2024), we use Ministral-8B-Instruct-2410, Mistral's newest and most powerful model under 10B.

The results of this process are 297,611 individual stance determinations (FAVOR, AGAINST, NONE). Outputs are encoded as both single-word strings and as numbers (1, -1, and 0) for plotting. Full model outputs for each stance are additionally logged for interpretability.

3.5 Sentiment Detection

Sentiment analysis is performed using two methods for comparison. We use the *Textblob* and *VADER* (Valence Aware Dictionary and sEntiment Reasoner) libraries.

Textblob – rule-based approach using pre-determined lexicon to assign scores, chosen for our data because this approach returns sentiment and subjectivity scores and is a good general tool for most document types. Sentiment returns scores from -1 to 1 and subjectivity scores from 0 to 1.

VADER – rule-based approach, specifically for social media texts, using pre-determined lexicon to assign scores, chosen for our data because this tool can capture emojis, slang and common social media word expressions. Sentiment scores are determined by a valence score from -4 to 4. The valence score includes positive, negative, neutral and compound scores with the compound score is the overall sentiment score.

We set the sentiment thresholds to be the same for both models, <-0.5 for negative, >0.5 for positive and used the compound score for the VADER model.

4 Analysis

4.1 Topic Distribution Analysis

Choose a topic: -- Select --

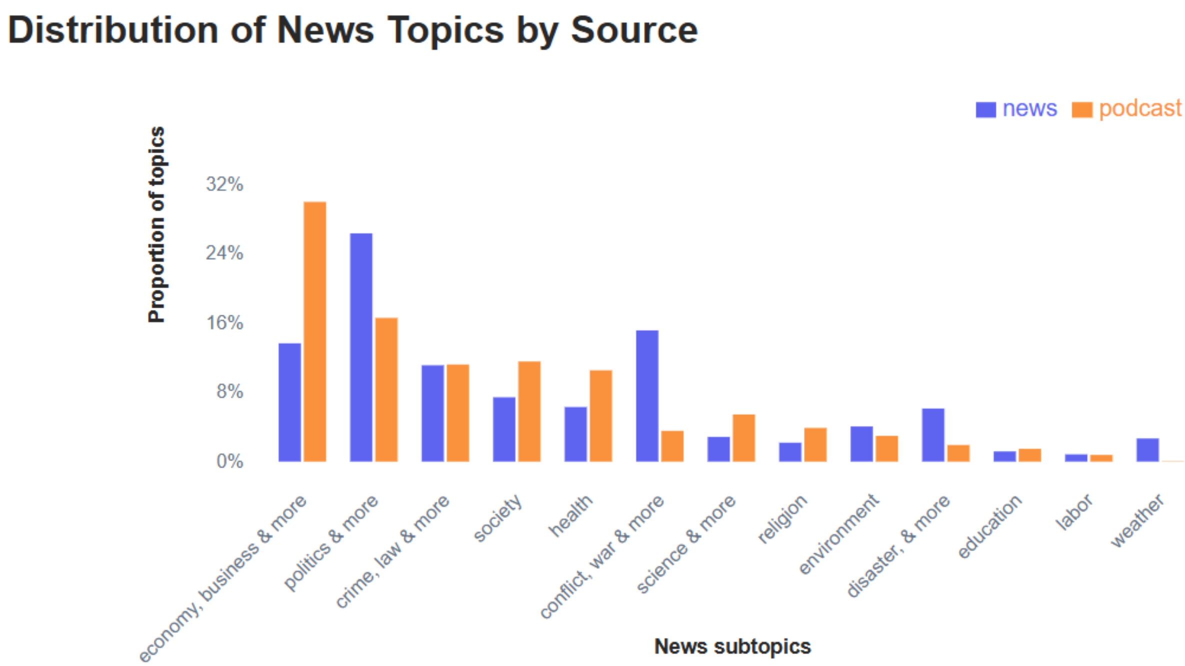

Figure 2: The above interactive paired bar chart shows IPTC topic distributions for key news topics across news and podcast mediums. The initial chart displays the highest-level topics, and the height of each bar represents the overall proportion of content dedicated to that topic by medium. Each topic can be drilled down to view a distribution of its constituent subtopics.

Overall, podcasts tend to under-cover hard news (e.g. war, public health, climate) in favor of personality driven politics, crime & security, and content related to culture and values. Public news, on the other hand, dedicates more space to coverage of institutions and policy.

War and conflict: 15% of public news vs 4% of podcasts.

Elections: Podcasts tend to cover candidates (68%) over procedures (8%). Public news allocates proportionally more coverage to process (32%). Podcasts focus on national races (13% vs 1% state), while public news is more balanced (6% national, 9% state).

Government and policy: Podcasts emphasize national security (65% vs 20%). Public news focuses on a greater range of policy issues like healthcare (15% vs 2%) and immigration policy (16% vs 8%).

Crime, law, and justice: Podcasts center on crime incidents (64% vs 43%), especially headline homicide cases (24% vs 10%). Public news features more routine judiciary coverage (40%), and a wider-range of crime issues like drug and cyber crime.

Society, values, rights: Podcasts give more space than public news to niche societal topics like men (16% vs. <1%), sexual behavior (49% vs. <8%), discrimination (7% vs 1%), and free speech/censorship.

Health and environment: Public news emphasizes communicable disease (69% vs 5%), while podcasts stress mental health (40%). Climate, resources, and sustainability topics are all more prominent in public news (42% vs 14%).

Family: In topics of family, news tends to cover family planning (72% vs 20%). Podcasts tend to cover dating & relationships (39% vs 2%).

To summarize, while both public news and podcasts contain substantial news content, podcasts seem to favor vivid incidents, personalities, and culture topics; public broadcasters cover a broader range of procedural, institutional, and policy centered topics.

4.2 Stance Analysis

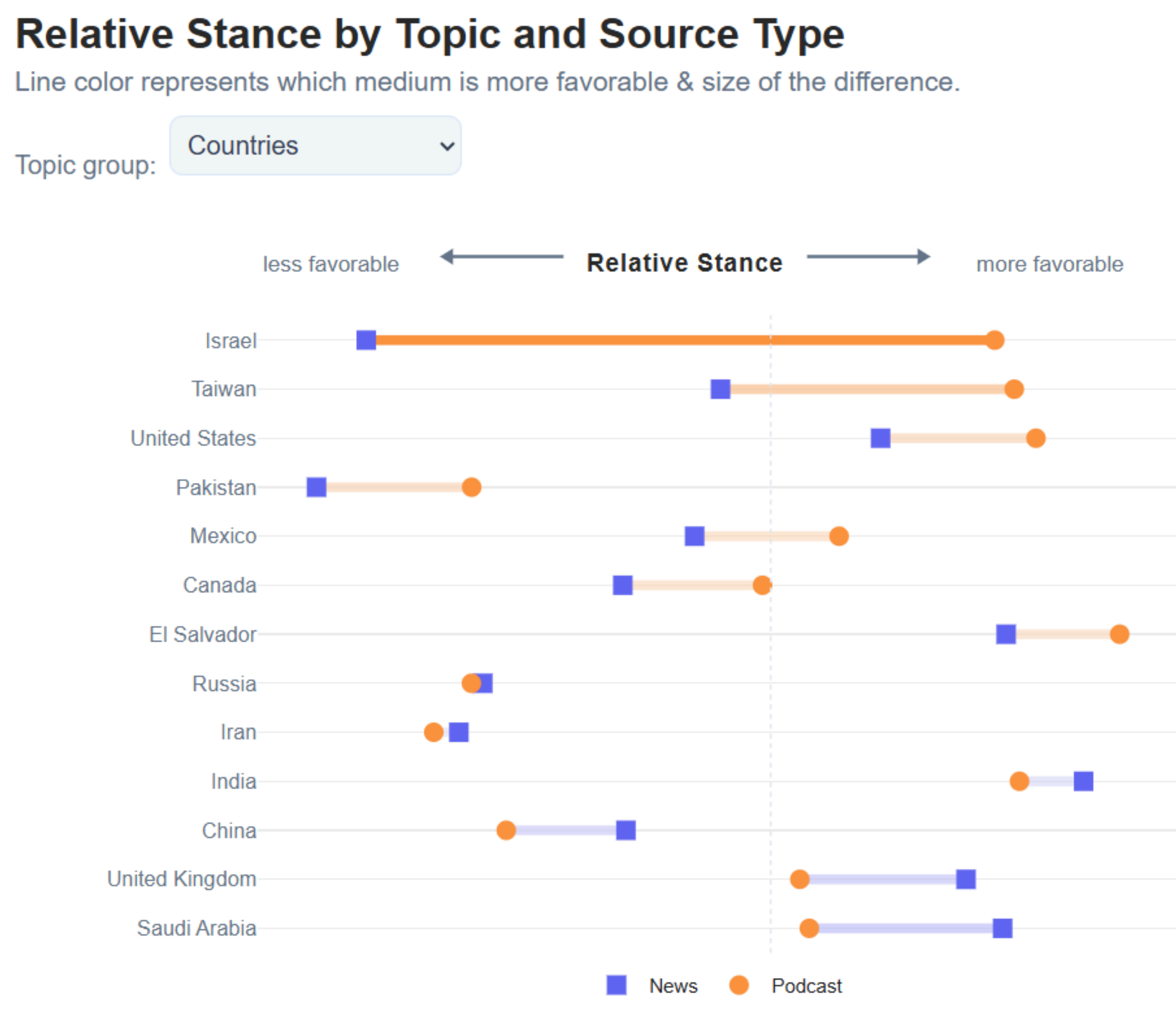

Figure 3: Interactive paired dot plot shows key news topics and relative stance scores (z-standardized) by medium. Dot color represents podcasts vs public news, and dot position represents relative stance. Line color indicates which source is more favorable, and line intensity shows the size of the gap. Rows are sorted in ascending order by Δ, where Δ is equal to mean_podcast minus mean_news (Δ > 0: podcasts more favorable).

We analyze three topic groups: **people**, **countries**, and **political issues**.

Note that reported stance scores use a FAVOR=1, NONE=0, and AGAINST=-1 encoding method. Reported correlations are weighted by topic Ns.

People

Alignment: r=0.74 (high). Both mediums cover basketball star Caitlin Clark most favorably, followed by the Pope. Sex offender Jeffrey Epstein is least favorable across both. Key political figures Donald Trump, Benjamin Netanyahu, and Elon Musk show minimal variation in average stance score by medium.

Bias: Compared to public news, podcasts tend, on average, to cover named public figures less favorably overall (negative mean Δ)

People covered *more* favorably in podcasts

- *The Trump cabinet* – Robert F. Kennedy Jr (Δ= +0.20), Pete Hegseth (+0.16).
- *Controversial figures* – Diddy (Δ= +0.30), Luigi Mangione (+0.17), Bob Menendez (+0.15), Kanye West (+0.15).

People covered *less* favorably in podcasts:

- *Democratic politicians* – Kamala Harris (Δ= −0.45), Nancy Pelosi (−0.47), Joe Biden (−0.38), Pete Buttigieg (−0.36), AOC (−0.33), Tim Walz (−0.33), Bernie Sanders (−0.34).
- *Foreign leaders* – Xi Jinping (Δ= −0.45), Mark Carney (−0.19), Vladimir Putin (−0.14), Volodymyr Zelensky (−0.16).

Countries

Alignment: r=0.56 (moderate). El Salvador and India have most favorable stance scores overall across mediums. Pakistan (news) and Iran (podcasts) are least favorable by source.

Bias: Slightly more favorable coverage of countries in podcasts overall.

Biggest differences: Israel (Δ = +0.29) is covered much more favorably in podcasts, as is Taiwan (Δ = +0.14).

Political Issues

Alignment: r=0.15 (weak – showing largest divergence). Despite this, nuclear power is the topic viewed most favorably across mediums, and average stances on tariffs, immigration, and police show minimal stance score variation.

Issues covered *more* favorably in podcasts: Racism (Δ = +0.31), nuclear power (+0.20), abortion (+0.17), climate change (+0.13), war (+0.10).

Note: for topics like “racism”, while a negative stance reflects opposition, “less negative” in podcasts does not necessarily imply support.

Issues covered *less* favorably in podcasts: Euthanasia (Δ = −0.53), capital punishment (−0.43), communism (−0.39), Democratic Party (−0.27), military service (−0.21).

4.3 Sentiment Analysis

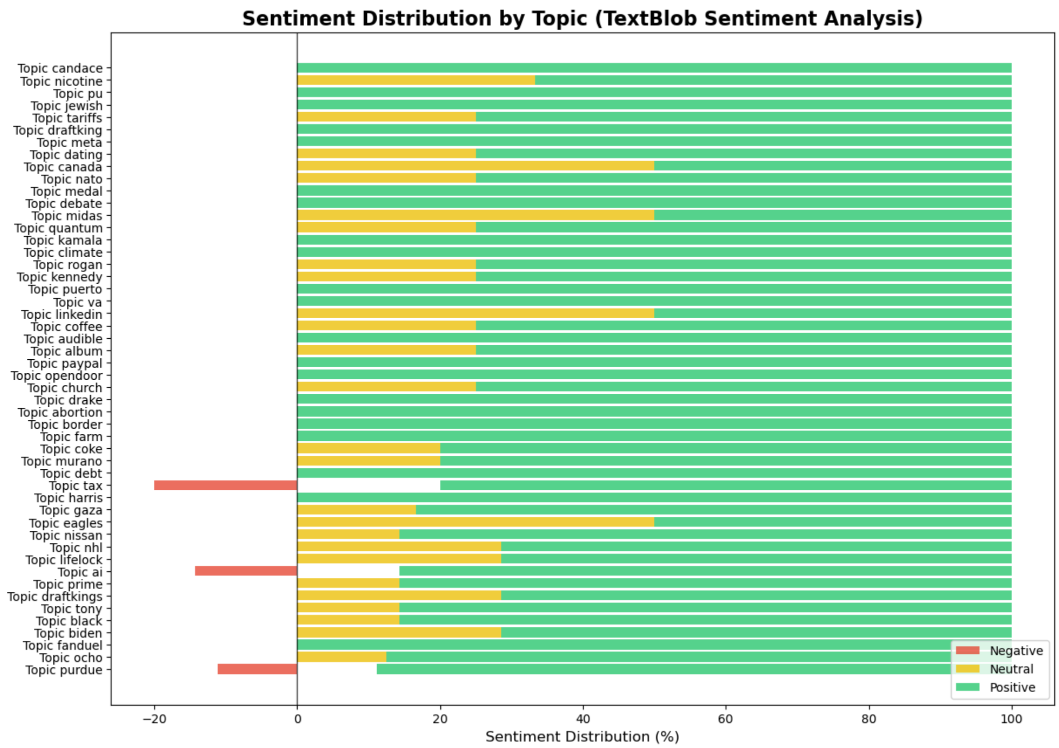

Figure 4: Textblob Sentiment Analysis results for top 50 topics using BERTopic results.

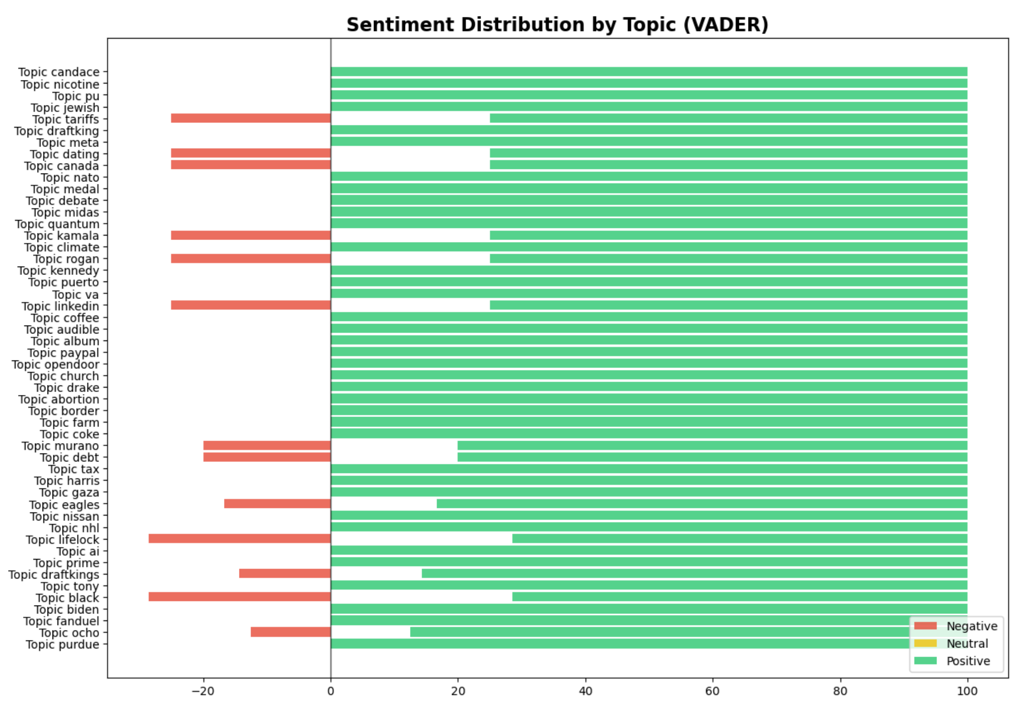

Figure 5: VADER Sentiment Analysis results for top 50 topics using BERTopic results

Textblob sentiment analysis produced neutral sentiment results, while VADER sentiment analysis only produced positive and negative sentiments among the top 50 topics. However, VADER analysis returned significantly more topics with negative sentiment. Overall, the data had a much higher percentage of positive sentiment than neutral or negative sentiment. The most polarized topics from the Textblob and VADER analysis are listed below.

Most Polarized Topics found in Sentiment Analysis of Top 50 Topics	
Textblob	VADER (Valence Aware Dictionary and sEntiment Reasoner)
Tax, AI, Purdue	*Tariffs, Dating, Canada, Kamala, Rogan, LinkedIn, Murano, Debt, Eagles, Lifelock, Draftkings, Black, Ocho*

Table 1 : Polarized topics based on sentiment analysis

Using the Textblob sentiment analysis, we also explored the subjectivity scores of our data. The subjectivity will return a score between 0 to 1 with scores closer to 0 suggesting more objective or factual data while scores closer to 1 suggest more subjective or opinion based data.

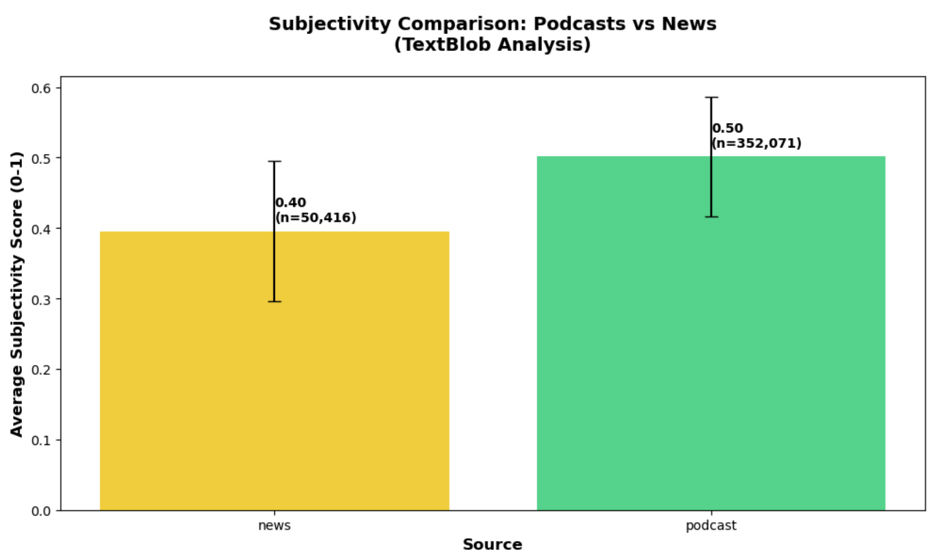

Figure 6: Average subjective score based on content source type (NPR,PBS news articles and podcasts)

In the above chart, we can see that the news article data has a lower subjectivity score than the podcast data. The lower score suggests that the news content is more factual than podcasts content, but still contains a level of subjectivity or opinion based information.

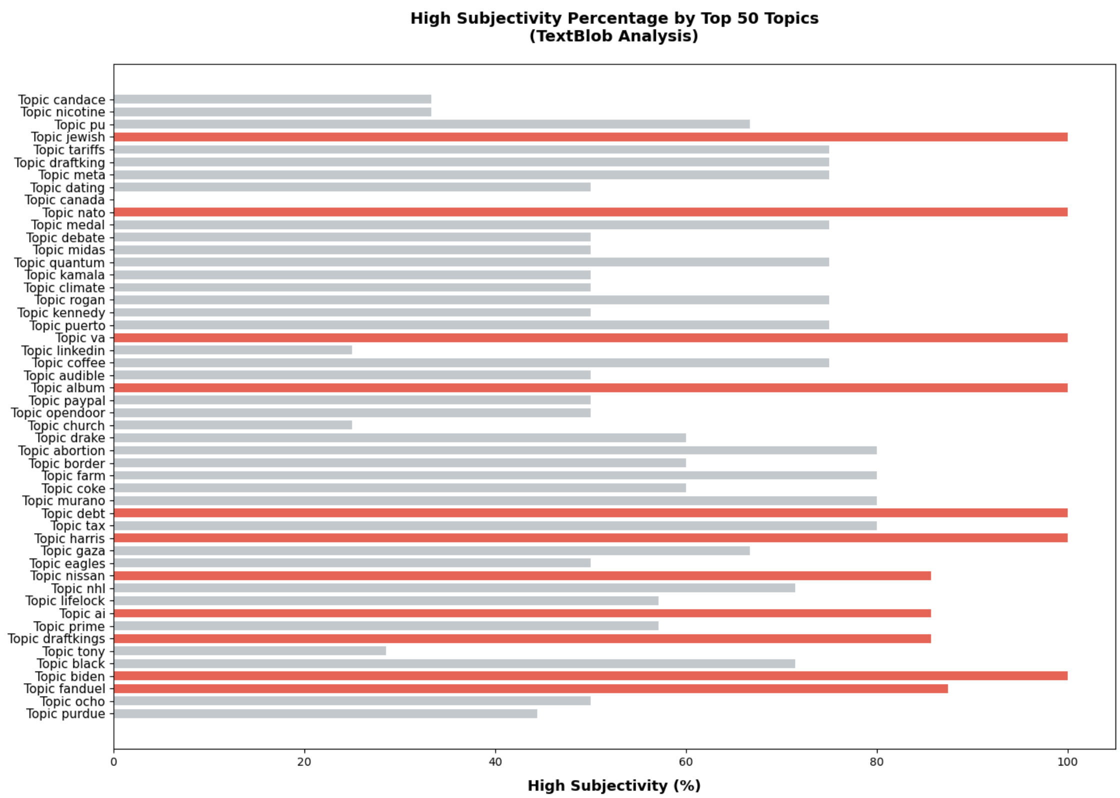

Figure 7: Subjectivity by top 50 topics

Figure 7 shows the percentage for each topic with high subjectivity. There are 11 topics, represented by red bars, where high subjectivity comprises greater than 85% of the content for the topic. The topics include: *Jewish, NATO, VA, Album, Debt, Harris, Nissan, AI, DraftKings, Biden* and *Fanduel*. Overall, we see that approximately half of the top 50 topics have high subjectivity scores comprising 75% or greater of content for the topic.

4.4 Framing Analysis

Figure 8: Interactive paired sentiment word cloud showing top 10 words by in-topic TF-IDF from two sources on a given topic. Color gradient represents sentiment (negative = red, neutral = gray, positive = green).

Differences in framing

Analysis of common words used across news topics shows that podcasts tend to emphasize *people* and lived *experiences* over news, which tends to emphasize *institutions* and *officials*. This is evidenced in top words: podcasts tilt to "*people, time, money, apple, school, women, work, years*", while news tilts to "*trump, president, federal, administration, government, state, health.*"

Which podcasts are *most* similar to news?

Agenda overlap: Of all podcasts, political podcasts have the highest agenda overlap with traditional news sources. Of NPR's 429 news topics, The Breakfast Club (310 shared topics), Ben Shapiro (307), Joe Rogan (295), Megyn Kelly (294), MeidasTouch (279) have the greatest alignment.

Lexical overlap: Political podcasts also tend to use the most similar vocabulary to traditional news as measured by median Jaccard similarity of top 20 TF-IDF words by topic. Top podcasts are Ben Shapiro (0.103), Pod Save America (0.096), MeidasTouch (0.094), Megyn Kelly (0.071).

Tone alignment: Tone alignment by per-topic sentiment correlation is moderate for left-leaning political shows (Pod Save America 0.355, MeidasTouch 0.361), but much lower for right leaning shows like Ben Shapiro (0.185). Comedy shows are the least aligned (Bad Friends −0.049; Smartless 0.069).

Consensus gaps

Overall, topics where many podcasts trend more positive than news skew towards lifestyle and behavior (e.g. sexual_behavior – 66% of podcasts more positive, mean Δ vs NPR +0.08).

Note: Most positive topics in podcasts are typically consumer / lifestyle topics (clothing, grocery, toys/games, streaming, health/beauty), suggesting noise in the data, some of which is likely sponsored advertiser content.

5 Discussion

5.1 Summary of Key Patterns

Over the course of a full year of content from podcasts and public news, we observe consistent systemic differences in the ways in which news is covered. Notably, podcasts allocate more space to personalities, vivid incidents, and discussion of culture and values. Public news focuses more on institutions, policy, and process. From stance analysis, we see that cross-medium alignment is strongest when discussing people, moderate when discussing countries, and weakest for political issues – a decrease as topics move from named individuals to more contested policy issues. Framing differences are reflected mainly in vocabulary, with podcasts describing more people and lived experiences, and public news centering more public officials and institutions.

5.2 Implications

For news audiences

- Choice of news medium matters, and can present the same news topics through different lenses. In particular, our finding of weak stance alignment for more contested political issues suggest that different mediums are likely to present diverging frames, even when tone and topic are similar.

For news producers

- Podcasts' interweaving of news topics with coverage of people, lived-experience, and culture show that while covering similar topics, they occupy the news landscape in very different ways. Traditional news still remains an anchor for "hard" process and policy content, and serves as a useful counterbalance for the more personal and incident-driven cycles of new media.

For researchers

- Results show the value a multifaceted approach (including topic modeling, sentiment analysis, stance detection, and framing) can bring to the understanding of a large text corpus.

5.3 Limitations

This study and its findings are exploratory in nature. We analyze only data from the calendar year July 1, 2024 – July 1, 2025, focus on only top U.S. podcasts and two public news outlets, and rely on largely automated labeling of topics, stance, sentiment, and framing. While we are careful to apply quality controls and manual review, several forms of noise and misclassification are likely still present (detailed in Section 6). We stress that results apply only to the podcasts and time period study, and that result magnitudes expressed in the paper are best interpreted as directional trends rather than exact point values.

6 Methodological Evaluation and Critique

6.1 Topic Modeling and Labeling

IPTC Media Topics

- **Advertiser noise:** Despite filtering for news-related topics and conducting manual review, advertising content, particularly from podcast ad-reads, sometimes slipped into our news topic data under economy or business subcategories.
- **Similar categories:** IPTC topics are assigned one-per-document. Some similar topics (i.e. "Immigration" vs "Immigration Policy") are subtopics of completely different major topics (i.e. "Society" vs "Governments and Politics"), and this may affect distributions.

Alternative string-matching labels

- **Heuristic matching:** Searching for special topics with heuristic matching (e.g., requiring "Trump" plus one of {Donald, President, Republican, Candidate}) reduces many false positives but is imperfect at reducing them all. Multi-word entities are especially error prone when searching through tokenized documents due to single-word acronyms colliding with other common words. Sentiment for US federal agency ICE, for instance, was spuriously positive due to podcast "ice-cold" beverage ads. This may be solved in future implementations with tools like spaCy PhraseMatcher.

6.3 Stance detection

- **Model-human agreement:** Reliability is moderate. In a 100-document audit, LLM labels matched human labels 66% of the time. FAVOR was the hardest class for the LLM to label correctly (recall = 0.36; F1 = 0.43).

- **Directional bias:** Hand labels were distributed NONE (52%), AGAINST (34%) and FAVOR (14%). The model, on the other hand, returned more AGAINST (47%) and fewer FAVOR (9%). Nearly half of mismatches followed a model=AGAINST and hand=NONE pattern. This suggests a tendency of the model to systematically read generalized criticism as being against the target topic (e.g. criticism of a specific Trump policy was read as being against the United States).

- **Source-specific bias:** Degree of agreement with hand labels varied by source, with higher alignment for more opinionated sources (e.g., NPR 0.44 vs The Ben Shapiro Show 0.75). This is likely due to negative directional bias when opinions are not present or are stated less clearly.

Note, however, that this is a limited sample-size, unstratified, single person audit. Further validation requires a larger sample stratified by source with multiple human raters.

- **Future improvements:** An additional step to Chain of Stance prompting could be prompting the model to identify where in the document the topic appears, and whether the evaluative language is directed at that topic. If the negativity is directed at another topic, default to NONE. This would be a relatively easy step to implement, and could potentially mitigate the biases encountered in this study.

6.4 Sentiment analysis

- **Context:** Even though we use two different sentiment analysis models – TextBlob (general) and VADER (social-media oriented) – both are rule-based, and can struggle with challenging tasks that involve sarcasm, negation, unusual slang, and topic-dependent context.

- **Chunking:** Using fixed thresholds of ±0.5 may over or underclassify certain sentiment labels.

6.5 Framing analysis

Noise reduction

- Use of TF-IDF to identify unique words can pick up on sponsors, numerics, and filler, especially in podcasts, reducing potentially useful lexical overlap. This could potentially be mitigated in future work by building a stoplist of tokens frequent in advertising or that are unique to the lexicon of certain podcasts. These can then be excluded before computing top TF-IDF words.

7 Statement of Work

7.1 Chelsea Simpson

- **Data acquisition:** scraped and processed all news items in the corpus.

- **Literature review:** surveyed related work, planned approaches, and compiled and formatted citations.

- **Sentiment analysis**: Implemented TextBlob and VADER pipelines, and processed by setting thresholds and computing polarity and subjectivity. Created sentiment visuals and synthesized key findings.

- **Topic labeling:** Contributed to the alternative named-entity topics. Generated the People/Countries/Political Issues lists and assisted with topic matching.

7.2 Sean Fontaine

- **Dataset acquisition:** Selected podcasts for study. Downloaded and processed podcast transcript data.

- **Topic modeling:** Configured and ran BERTopic. Matched and verified found topics with closest IPTC topic label.

- **Stance detection:** Implemented Chain of Stance prompting. Ran Ministral-8B-Instruct-2410 on AWS and critically evaluated results.

- **Framing analysis:** Ran TF-IDF across all sources for all topics to identify key framing terms. Identified key framing patterns.

- **Interactive visuals:** Created front and backend structures to host interactive visuals and blog.

References

Corporation for Public Broadcasting. (n.d.). *About CPB*. Retrieved August 14, 2025, from https://cpb.org/aboutcpb

DeMets, S., & Spiro, E. (2025). Podcasts in the periphery: Tracing guest trajectories in political podcasts. *Social Networks*, *82*, 65–79. https://doi.org/10.1016/j.socnet.2025.01.001

Edison Research. (2025, May 2). *The top 50 podcasts in the U.S. for Q1 2025 from Edison Podcast Metrics*. https://www.edisonresearch.com/the-top-50-podcasts-in-the-u-s-for-q1-2025-from-edison-podcast-metrics/

Ferret, O., Déjean, H., & Nioche, J. (2016). Knowledge-based Information Extraction from a Large-scale Text Corpus for Geopolitical Crisis Assessment. In *Proceedings of the 13th Conference on Natural Language Processing (KONVENS 2016)* (pp. 65–71). Bochumer Linguistische Arbeitsberichte. https://aclanthology.org/W16-1306.pdf

Google. (n.d.). *Speech-to-Text: AI speech recognition and transcription*. Google Cloud. Retrieved August 14, 2025, from https://cloud.google.com/speech-to-text

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure* (arXiv:2203.05794). arXiv. https://arxiv.org/abs/2203.05794

Hutto, C. J., & Gilbert, E. E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the Eighth International Conference on Weblogs and Social Media (ICWSM-14)*. The AAAI Press.

International Press Telecommunications Council. (2025, August 13). *Media Topics (en-US)- IPTC NewsCodes controlled vocabularies*. https://www.iptc.org/std/NewsCodes/treeview/mediatopic/mediatopic-en-US.html

Loria, S. (2018). *TextBlob documentation (Release 0.15.2)*. https://textblob.readthedocs.io/en/dev/

Ma, J., Wang, C., Xing, H., Zhao, D., & Zhang, Y. (2024). *Chain of Stance: Stance Detection with Large Language Models* (arXiv:2408.04649). arXiv. https://arxiv.org/abs/2408.04649

Mistral AI. (2025). *Ministral-8B-Instruct-2410* [Large language model]. Hugging Face. https://huggingface.co/mistralai/Ministral-8B-Instruct-2410

Newman, N. (2025, June 17). *Overview and key findings of the 2025 Digital News Report*. Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2025/dnr-executive-summary

NPR. (2025). *News*. Retrieved August 14, 2025, from https://www.npr.org/sections/news/

PBS. (2025). *PBS News*. Retrieved August 14, 2025, from https://www.pbs.org/newshour/

Pew Research Center. (2023, April 18). *Podcasts as a source of news and information*. https://www.pewresearch.org/journalism/2023/04/18/podcasts-as-a-source-of-news-and-information/

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825–2830.

PodScribe. (n.d.). *Transcripts of the most popular podcasts*. https://podscribe.app/

Rudnik, C., Ehrhart, T., Ferret, O., Teyssou, D., Troncy, R., & Tannier, X. (2019). Searching news articles using an event knowledge graph leveraged by Wikidata. In *Companion Proceedings of the 2019 World Wide Web Conference (WWW '19 Companion)* (pp. 957-964). ACM. https://doi.org/10.1145/3308560.3316761

Yu, L., & Yang, L. (2024). News media in crisis: A sentiment and emotion analysis of US news articles on unemployment in the COVID-19 pandemic. *Humanities and Social Sciences Communications*, *11*(1), Article 854. https://doi.org/10.1057/s41599-024-03225-9

Appendix A: Corpus Metadata

Source Name	Documents
NPR	29,425
PBS	20,919
Armchair Expert With Dax Shepard	170
Bad Friends	101
Call Her Daddy	97
Cancelled With Tana Mongeau Brooke Schofield	35
Candace	210
Club Shay Shay	1,171
Conan O'Brien Needs A Friend	112
Crime Junkie	85
Distractible	101
Huberman Lab	85
Impaulsive With Logan Paul	41
Kill Tony	54
Matt And Shane's Secret Podcast	64
Million Dollaz Worth Of Game	52
Morbid	111
Mrballen Podcast Strange Dark Mysterious Stories	117
Murder Mystery Makeup	40
My Favorite Murder With Karen Kilgariff And Georgia Hardstark	155
New Heights With Jason Travis Kelce	54
Pardon My Take	150
Pod Save America	154
Rotten Mango	77
Shawn Ryan Show	98
Smartless	69
Stuff You Should Know	215
Talk Tuah With Haliey Welch	15
Ted Talks Daily	363
The Ben Shapiro Show	321
The Bill Simmons Podcast	143
The Breakfast Club	979
The Joe Budden Podcast	97
The Joe Rogan Experience	181
The Lol Podcast	107
The Megyn Kelly Show	382
The Meidastouch Podcast	1,219
The Mel Robbins Podcast	113
The Ramsey Show	260
The Tucker Carlson Show	147
This Past Weekend w Theo Von	79

Vince	320

Appendix B: Chain of Stance Prompting Format

“[TASK]

You are an expert in stance detection.

Your task is to determine the stance of a given text towards a specific topic.

Follow these steps carefully to provide a complete analysis and a final conclusion.

Source Name: "{source_name}"

Title: "{title}"

Text for Analysis: "{text}"

Topic: "{topic}"

Step 1: Contextual Information Analysis

Analyze the contextual information of the text.

Consider the topic, the likely identity of the author, the target audience, and any relevant socio-cultural background.

Step 2: Main Idea and Viewpoint Identification

Based on the text and context, what are the core viewpoints and main intentions being expressed regarding the topic?

Step 3: Language and Emotional Attitude Analysis

Analyze the language, tone, and emotion.

Identify emotive words, rhetorical devices, and the author's overall tone (e.g., affirmative, negative, neutral, sarcastic).

Step 4: Comparison with Possible Stances

Compare the text's content and tone against the three possible stances (FAVOR, AGAINST, NONE).

For each stance, list evidence from the source (if any) of that stance.

Step 5: Logical Inference and Consistency Check

Synthesize your analysis from all previous steps to make a final decision on the most likely stance expressed in the text from (FAVOR, AGAINST, NONE).

Step 6: Final Stance Determination

Output the final stance on a new line, in the format 'Final Stance: [STANCE]', where [STANCE] is one of FAVOR, AGAINST, or NONE.

Begin your analysis now.

[/TASK]”

Appendix C: Topics Used for Stance Detection

People

- Alexandria Ocasio-Cortez
- Benjamin Netanyahu
- Bernie Sanders
- Bob Menendez
- Caitlin Clark
- Chuck Schumer
- Diddy
- Donald Trump
- Elon Musk
- JD Vance
- Jeff Bezos
- Jeffrey Epstein
- Joe Biden
- Justin Trudeau
- Kamala Harris
- Kanye West
- Kevin McCarthy
- Luigi Mangione
- Mark Carney
- Mark Zuckerberg
- Mike Johnson
- Mitch McConnell
- Nancy Pelosi
- Pete Buttigieg
- Pete Hegseth
- Pope
- Robert F. Kennedy Jr.
- Ron DeSantis
- Sam Altman
- Taylor Swift

Countries

- Canada
- China
- El Salvador
- India
- Iran
- Israel
- Mexico
- Pakistan
- Russia
- Saudi Arabia
- Taiwan
- United Kingdom
- United States

Political Issues

- Abortion
- Capital Punishment
- Christian Orthodoxy
- Civil Rights
- Climate Change
- Communism
- COVID-19
- Crypto
- Democratic Party
- Dictatorship
- Discrimination
- Diversity, Equity and Inclusion
- Environmental Policy
- Euthanasia
- Family Planning
- FEMA
- Genocide
- Global Warming
- Government Aid
- ICE
- Immigration
- Immigration Policy
- India–Pakistan
- Israel–Gaza
- LGBTQ
- Military Service
- Nuclear Policy
- Nuclear Power
- Opioids
- Personal Weapon Control Policy

- Taylor Swift–Travis Kelce
- Tim Cook
- Tim Walz
- Vladimir Putin
- Volodymyr Zelensky
- Xi Jinping
- Police
- Pornography
- Racism
- Religion
- Republican Party
- Russia–Ukraine
- Tariffs
- Terrorism
- Texas Floods
- TikTok
- Tobacco and Nicotine
- Unions
- USAID
- Vaccine
- War
- Welfare