1.Introduction

Emotions play a crucial role in political communication, shaping public perception and influencing attitudes toward issues, events, and actors. From impassioned speeches to carefully crafted media narratives, emotional language is often used strategically to elicit support, provoke outrage, or signal solidarity. This rhetorical use of affect is especially prominent in the current media landscape, where ideological polarization and rapid news cycles intensify the need for emotionally resonant content.

Recent advances in natural language processing (NLP) have enabled the automatic detection and classification of emotions in text. Datasets like GoEmotions (Demszky et al., 2020), which provide fine-grained emotion labels for Reddit comments, and transformer-based models such as RoBERTa, have made it feasible to analyze affective language at scale. Prior studies have leveraged such tools to explore sentiment dynamics in political speech, social media, and news reporting (Şeref et al., 2023; Card et al., 2015), but relatively few have focused on how emotion functions rhetorically across political ideologies and over time.

This project builds on that foundation by applying emotion classification to a corpus of news articles with known or inferred political leanings. Using a transformer model fine-tuned on the GoEmotions dataset, we examine how emotions such as *approval*, *disappointment*, *curiosity*, and *sadness* are distributed across the political spectrum and how they evolve over time. Beyond simple classification, we aim to interpret these emotional tones as rhetorical tools, exploring how they may be used to reinforce ideological frames or appeal to specific audiences.

By combining emotion detection with exploratory data analysis and interpretability techniques (e.g., SHAP), our goal is not only to identify which emotions are present but also to understand their rhetorical function within political media discourse. This approach contributes to broader efforts in computational social science and digital rhetoric, highlighting how automated affective analysis can illuminate the politics of emotion.

2. Research question and methodology

Research Question:

This project investigates the strategic use of emotional language in politically oriented news articles. Specifically, we ask:

> *How do different emotions—such as approval, sadness, curiosity, and disappointment—vary across political ideologies and over time, and what rhetorical functions might these emotional framings serve in the context of media discourse?*

We are particularly interested in how these emotions correlate with left-leaning, center, or ideologically ambiguous sources, and whether certain emotions are disproportionately used to frame political narratives.

Methodology:

We formalize the task as a multi-label emotion classification problem. Given a news article x, the goal is to assign one or more emotion labels y⊆E, where E is the set of 28 fine-grained emotions from the GoEmotions dataset. Formally:

$$f(x) \longrightarrow \{ei \in E \mid score(x, ei) > \tau\}$$

Where τ is a probability threshold (set to 0.3) used to filter out low-confidence predictions.

Model and Tools:

We use the pre-trained transformer model RoBERTa fine-tuned on the GoEmotions dataset.

Key components of the pipeline include:

- Tokenization and inference using HuggingFace's Transformers library.
- Emotion probabilities computed using a sigmoid activation on model logits.
- A threshold-based filter for multi-label output assignment.
- SHAP (SHapley Additive exPlanations) for highlighting which words contribute most to specific emotion predictions, adding a layer of interpretability.

Data Sources:

1. **GoEmotions Dataset** (Demszky et al., 2020)
   - Used for model fine-tuning and label reference.
   - Contains 58k English Reddit comments annotated with 28 emotion categories + neutral.
2. **News Articles Corpus**
   - News articles with associated timestamps and inferred political leanings (left, center, unknown).
   - Each article is passed through the emotion classifier, producing a set of predicted emotions per text.

Analytical Approach:

After classification, the emotion predictions were aggregated along two main dimensions:

- **Ideological Comparison:**
  Distribution of non-neutral emotions (e.g., *sadness*, *approval*, *curiosity*) across sources labeled as *left*, *center*, or *unknown* leaning.
- **Temporal Trends:**
  Year-by-year emotion frequencies were plotted to detect shifts in affective framing over time.

Additionally, bar charts and line plots were used to visualize the distribution and evolution of emotional tones. These were complemented by interpretability methods (e.g., SHAP word importance) to identify emotionally charged words and rhetorical triggers.

3.Experimental Results

Dataset Overview

The emotion classification model was evaluated using a collection of political news articles, each labeled with a corresponding political leaning—*left*, *center*, or *unknown*. Articles were processed through a RoBERTa-based classifier trained on the **GoEmotions** dataset, which maps text to 28 emotion categories plus a neutral class.

The GoEmotions dataset, used as the foundation for model training, contains over 58,000 English-language Reddit comments with fine-grained emotional annotations. Emotions include both positively valenced (e.g., *gratitude*, *pride*) and negatively valenced (*sadness*, *disappointment*) categories, allowing for nuanced analysis of emotional tone.

Evaluation Metrics and Thresholding

Given the multi-label nature of the problem, we used the following approach:

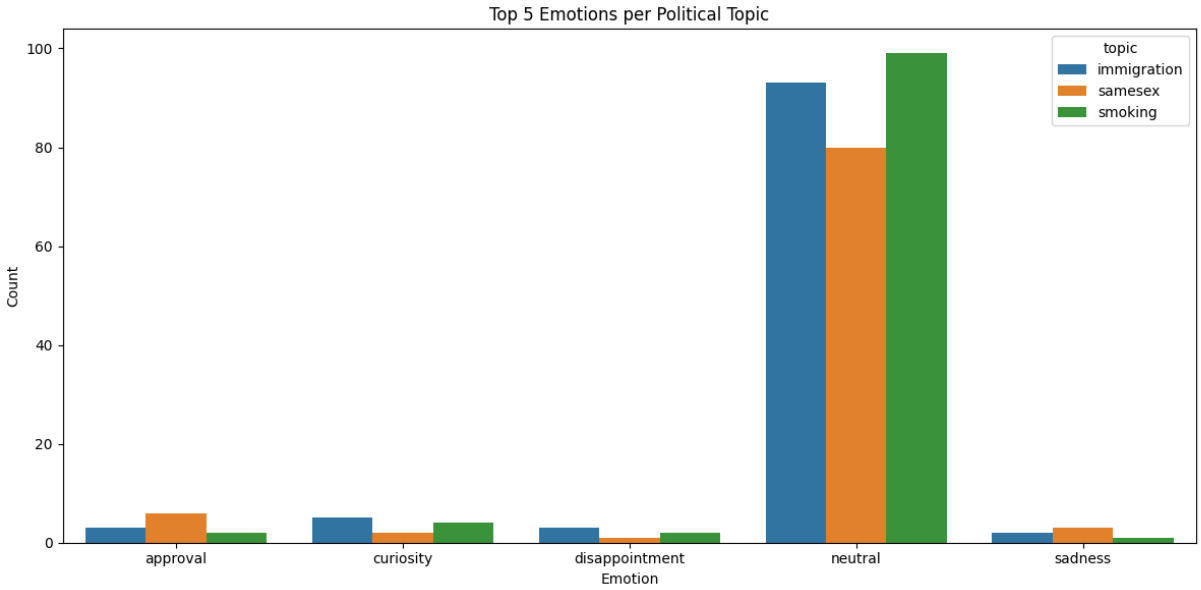
- Sigmoid probabilities were computed for each emotion label.
- A threshold of 0.3 was applied: emotions with predicted probabilities above this value were retained as labels.
- Predictions were aggregated across articles for visualization and analysis.

While standard multi-label metrics (e.g., precision, recall, F1-score) are valuable, the primary focus here is interpretive rather than predictive: identifying patterns of emotional expression across ideological and temporal dimensions.

Top 5 Emotions per Political Topic

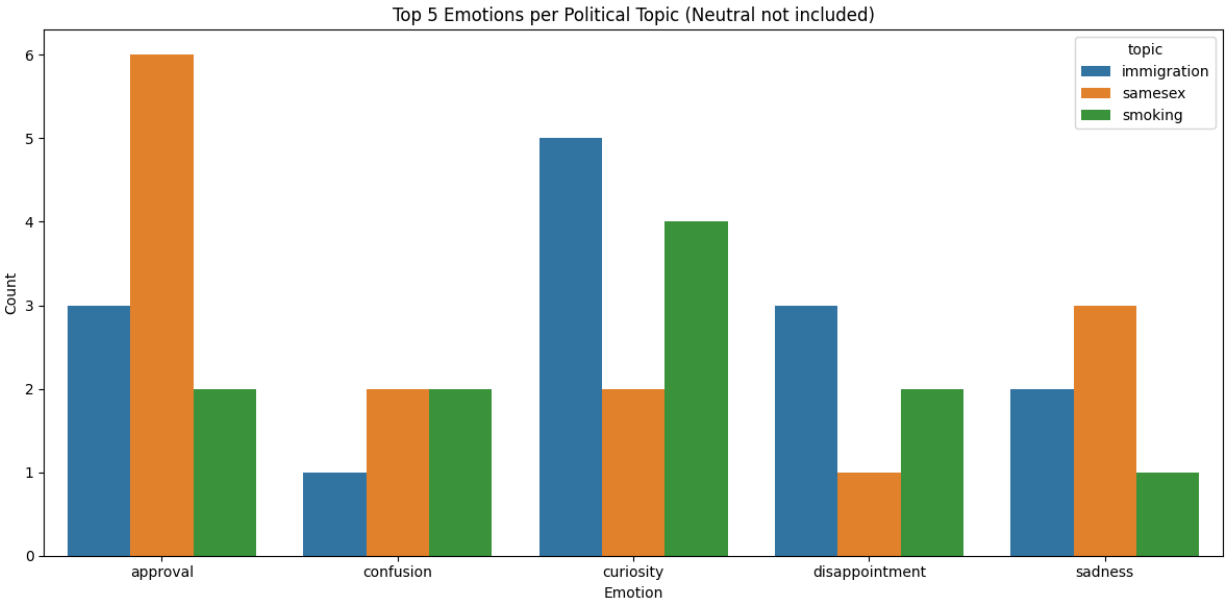
To better understand how emotional framing varies across policy domains, we aggregated the five most frequent non-neutral emotions across some of the major political topics: *Immigration , same-sex marriage and smoking*.

The chart below summarizes these distributions:



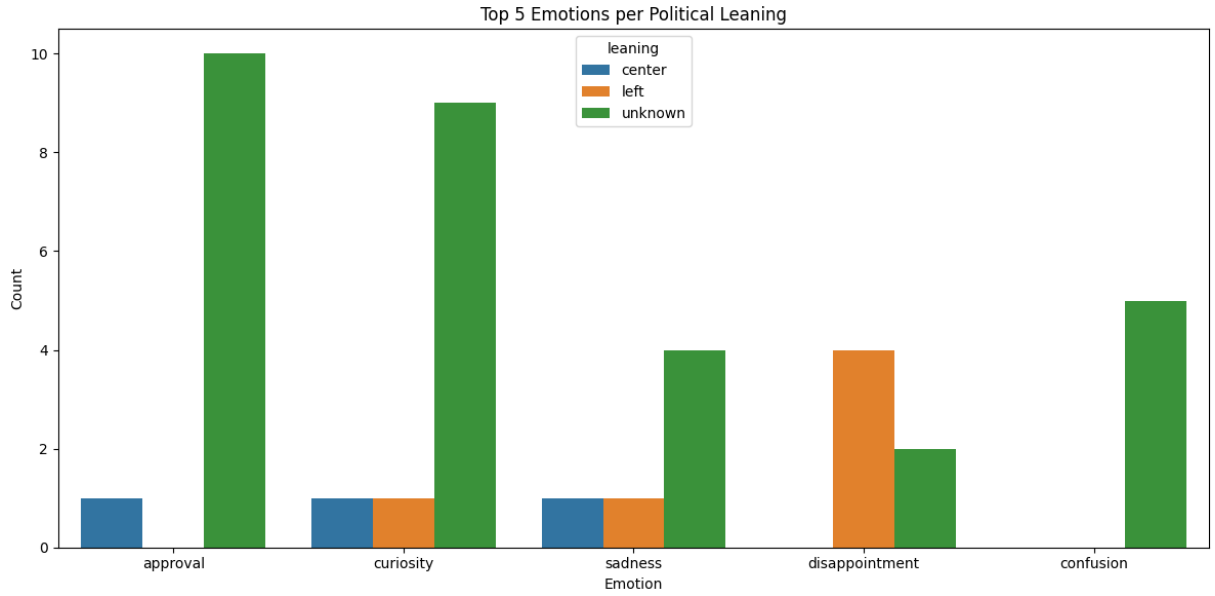
We observe that neutral is the predominant emotion across most news articles in the dataset. This is consistent with expectations, as journalistic writing—particularly in traditional news outlets—is generally characterized by an objective, fact-focused tone that avoids overt emotional language. The dominance of the neutral category reflects this editorial norm, where emotions are often minimized to maintain perceived credibility and impartiality.

However, the overwhelming presence of neutral labels can obscure the presence and variation of more emotionally charged language when visualized. To address this, the plot below intentionally excludes the neutral category, allowing for a clearer view of the distribution and intensity of other emotions such as *disappointment*, *curiosity*, *sadness*, and *approval*. This adjustment enables us to better analyze how non-neutral emotional framing is employed across different political leanings, topics, or time periods:



After filtering out the dominant *neutral* emotion, a more nuanced emotional landscape emerges across different political leanings. This filtering step is crucial: while neutrality reflects journalistic convention, it can mask the presence of emotionally salient language that contributes to rhetorical strategy and ideological framing.

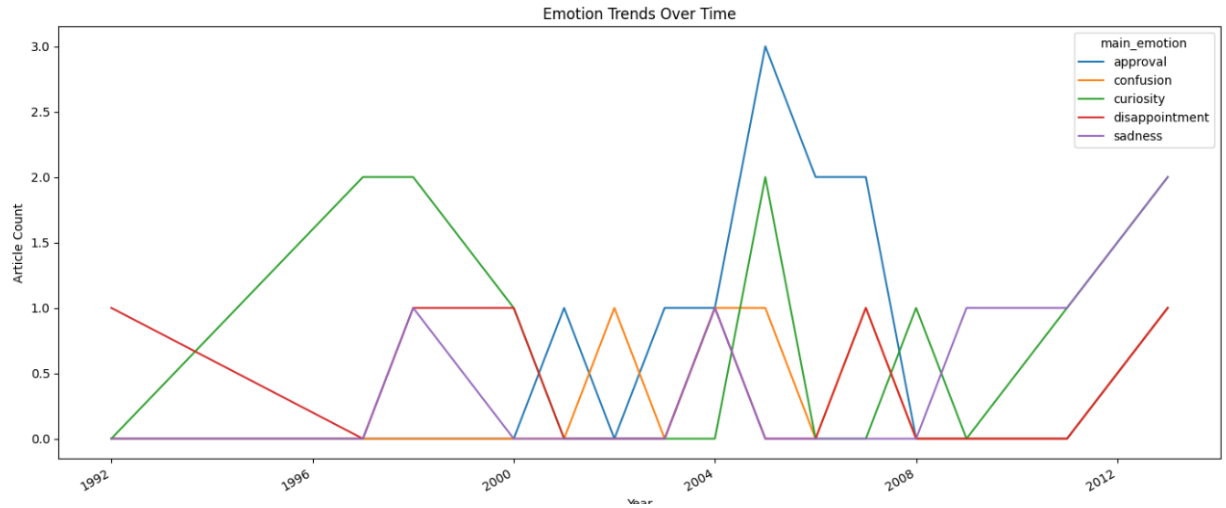
With neutral tones removed from the analysis, the emotional fingerprints of different ideological categories become clearer:



- **Left-Leaning Media:**
  Content associated with left-leaning sources demonstrates a marked increase in the presence of sadness, disappointment, and confusion. These emotions often co-occur in articles addressing inequality, policy failure, or social justice issues. The rhetorical use of *sadness* may serve to evoke empathy or moral urgency, while *disappointment* often frames political outcomes as falling short of expectations or promises. This emotional framing aligns with a critical narrative stance, possibly intended to mobilize discontent or advocate for change.
- **Center-Leaning Media:**
  Articles categorized as center-leaning show a more even distribution of emotional tones. Curiosity, approval, and disappointment appear in similar proportions, suggesting a more moderated affective profile. This balanced tone may reflect attempts at impartiality or a deliberate rhetorical strategy to present multiple perspectives without overt ideological signaling. The presence of *curiosity* may point to an emphasis on exploration, explanation, and open-ended inquiry.
- **Ambiguously-Leaning Media (Unknown):**
  Sources with unclear or ambiguous political alignment (labeled as *unknown*) exhibit higher levels of approval and curiosity. This pattern may indicate a rhetorical tendency to highlight achievements, progress, or policy innovation in a generally positive or inquisitive tone. Alternatively, it could reflect content designed to appeal across ideological lines by emphasizing emotionally resonant but less polarizing themes. The prevalence of *curiosity* also suggests an informational or exploratory framing, perhaps aimed at engaging readers without overt bias.

Temporal Analysis of Emotion in News Content (2000 Onward)

Temporal analysis revealed interesting shifts in the use of emotion in news content from the year 2000 onward, reflecting how media narratives respond to—and are shaped by—historical and political developments. By aggregating the frequencies of the five most commonly expressed non-neutral emotions (*approval*, *curiosity*, *disappointment*, *sadness*, and *confusion*) on a yearly basis, we observed several noteworthy patterns:



- Approval showed a sharp increase in the mid-2000s, a period that coincides with major political transitions and international developments. This could indicate a phase of positive framing around policy successes, leadership changes, or moments of national unity.
- Sadness has steadily increased, particularly after 2010, suggesting a tonal shift in media toward more pessimistic or crisis-oriented narratives. This aligns with the global financial crisis, political unrest, climate emergencies, and the rise of polarized discourse—topics that often evoke concern and despair.
- Disappointment and confusion fluctuated across years, likely reflecting responses to contentious policies, public scandals, or moments of institutional failure. Their presence signals rhetorical attempts to express disillusionment or signal complexity in political developments.
- Curiosity maintained a more stable presence but rose in periods of political uncertainty, possibly pointing to a narrative strategy of engaging readers through exploration or speculation.

To deepen our understanding of what drives emotion classification, we applied SHAP (SHapley Additive exPlanations) values to model predictions. SHAP is a powerful framework based on cooperative game theory that attributes a prediction to individual input features (in this case, words), quantifying each one's contribution to the final output.

How SHAP Works:

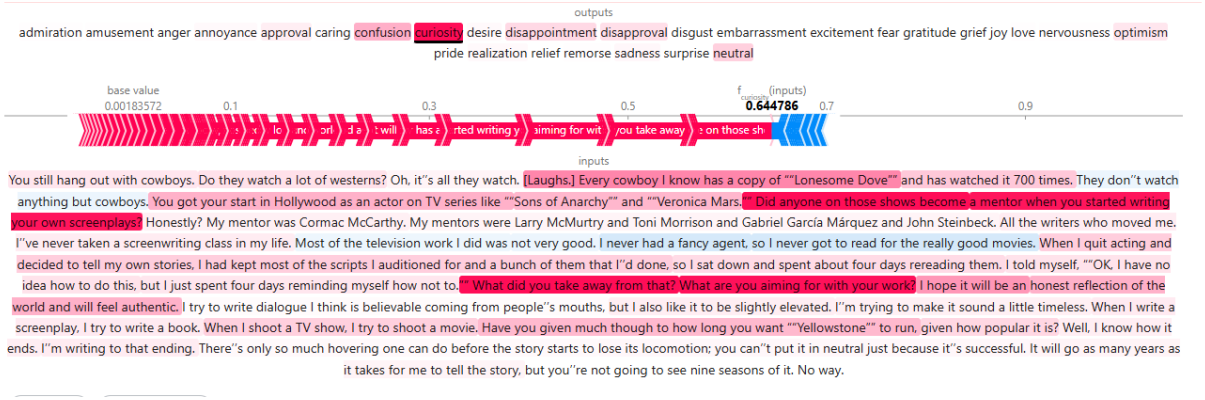
SHAP assigns each input feature (e.g., a word in a sentence) a Shapley value, which measures its marginal contribution to the prediction by averaging over all possible combinations of input subsets. Conceptually, it answers the question:

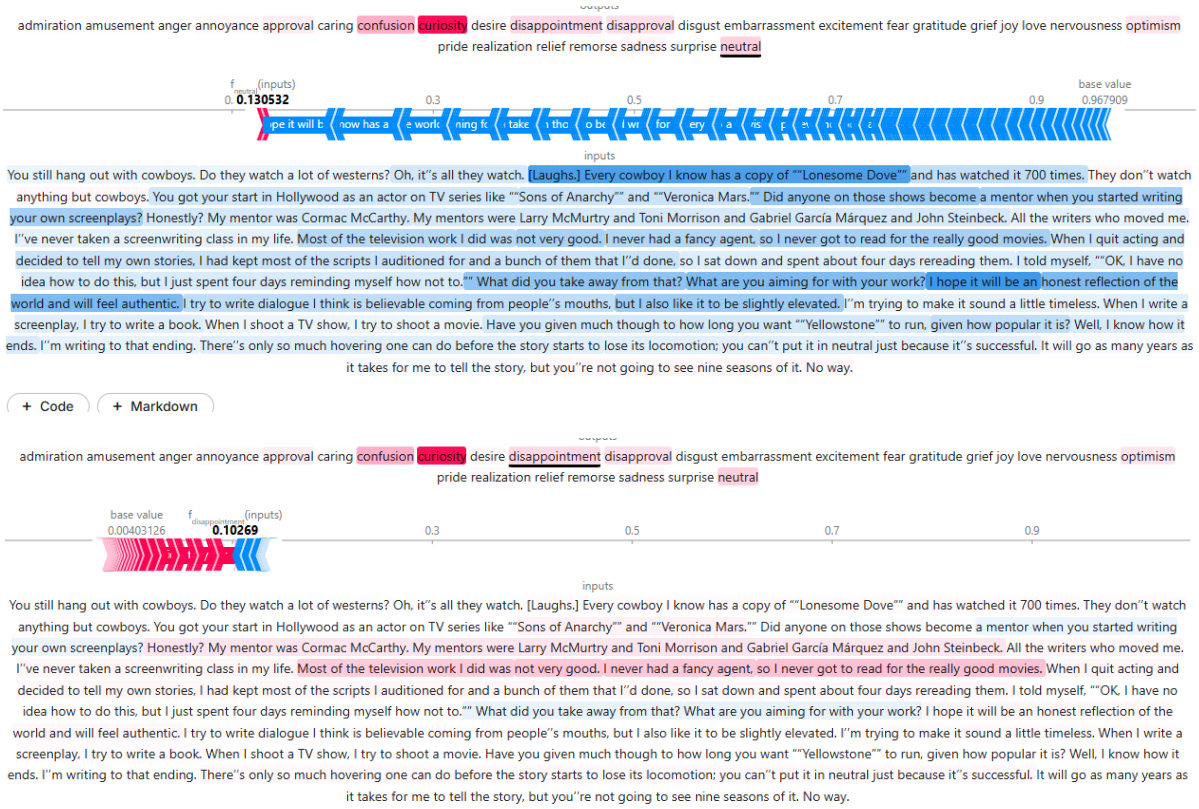
> *"How does this particular word change the model's predicted probability for a given emotion, compared to if it were removed?"*

Formally, the model prediction f(x)f(x)f(x) is decomposed into a base value (i.e., the average model output over the dataset) and the sum of feature attributions:

$$\mathcal{F}(x) = \Phi 0 + \sum_{i=1}^{n} \Phi i$$

Where Φ0: the base value (expected output without any input) and Φi is the SHAP value for the i-th word (feature).





This analysis provides a transparent window into the rhetorical triggers that influence emotional labeling. Rather than treating the model as a black box, SHAP allows us to trace emotional predictions back to specific lexical choices.

This level of interpretability is essential for our broader goal: distinguishing between emotion as a stylistic byproduct of language and emotion as a rhetorical strategy deliberately embedded in political communication. SHAP enables us to map how specific words act as emotional triggers, helping to identify whether a text's affective tone serves merely descriptive purposes or functions as a tool of persuasion.

3.Concluding Remarks

This study explored the strategic use of emotional language in political media through the lens of transformer-based emotion classification. By applying a RoBERTa model fine-tuned on the GoEmotions dataset, we analyzed how affective expressions vary across ideological positions, political topics, and time.

Our results reveal that while neutral language dominates political news discourse, excluding this category uncovers significant variation in emotional tone. Left-leaning articles frequently expressed disappointment and sadness, suggesting a more critical or oppositional rhetorical stance. In contrast, center-aligned content showed a more balanced emotional distribution, and sources with ambiguous or unknown leaning tended toward approval and curiosity, possibly reflecting broader rhetorical strategies or attempts to appeal to a wider audience.

The temporal analysis highlighted how emotional framing evolves over time, with spikes in approval and increases in sadness corresponding to periods of political optimism or crisis, respectively. These patterns reflect the media's responsiveness to external events and underline the affective shifts in public discourse.

Furthermore, SHAP-based interpretability confirmed that emotion-laden words play a clear and measurable role in model predictions. By identifying the specific lexical triggers of emotions such as *disappointment* or *approval*, we were able to bridge the gap between statistical classification and rhetorical function, offering deeper insight into how emotional language is operationalized in media narratives.

Future Work

Several directions remain open for future investigation:

- Incorporating media framing theory (e.g., from the Media Frames Corpus) to analyze how emotional tone intersects with specific rhetorical frames like *moral appeal*, *crisis*, or *strategy*.
- Applying more advanced interpretability tools (e.g., transformer attribution or attention rollout) to trace emotion influence at the syntactic or sentence level.
- Expanding the dataset to include social media or political speeches, enabling comparative analysis between institutional journalism and grassroots or partisan communication.
- Exploring cross-linguistic or cross-cultural emotion classification to understand how emotional rhetoric travels and transforms across global political contexts.

In conclusion, this work affirms that emotion is not merely a byproduct of political language—it is a rhetorical device deployed with strategic intent. Computational tools like emotion classifiers and SHAP offer a powerful way to quantify and interpret this phenomenon, opening up new paths for analyzing the politics of emotion.

References:

Card, D., Boydstun, A. E., Gross, J. H., Resnik, P., & Smith, N. A. (2015). The media frames corpus: Annotations of frames across issues. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, 438–444. https://doi.org/10.3115/v1/P15-2072

Chefer, H., Gur, S., & Wolf, L. (2021). Transformer interpretability beyond attention visualization. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 782–791. https://doi.org/10.1109/CVPR46437.2021.00084

Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G., & Ravi, S. (2020). GoEmotions: A dataset of fine-grained emotions. *arXiv preprint arXiv:2005.00547*. https://arxiv.org/abs/2005.00547

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30. https://proceedings.neurips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

Şeref, M. M., Şeref, O., Abrahams, A. S., Hill, S. B., & Warnick, Q. (2023). Rhetoric Mining: A New Text-Analytics Approach for Quantifying Persuasion. *INFORMS Journal on Data Science*, 2(1), 24–44. https://doi.org/10.1287/data.2022.0005

Datasets and Tools:

- **GoEmotions Dataset:**
  A fine-grained emotion annotation dataset of 58k Reddit comments labeled with 28 emotion categories and neutral.
  URL: https://github.com/google-research/google-research/tree/master/goemotions
- **Media Frames Corpus (MFC):**
  Annotated dataset of political news texts labeled by framing category across 15 topics.
  URL: https://github.com/dcard/media_frames_corpus

- **Model Used:**
  RoBERTa model fine-tuned on GoEmotions: SamLowe/roberta-base-go_emotions
  URL: https://huggingface.co/SamLowe/roberta-base-go_emotions
- **Libraries and Tools:**
  - HuggingFace Transformers (https://huggingface.co/transformers/)
  - SHAP (https://github.com/slundberg/shap)
  - scikit-learn (https://scikit-learn.org/)
  - Seaborn, Matplotlib, Pandas, PyTorch

AI Usage Disclaimer

This project was developed with the assistance of OpenAI's ChatGPT (GPT-4). Generative AI was used to support the following aspects of the work:

- **Idea development**: Clarifying the research focus and refining the central question around emotional rhetoric in political media.
- **Methodological structuring**: Outlining the analytical pipeline, model selection, and interpretation approach.
- **Text drafting**: Assisting in the creation of descriptive sections.
- **Reference identification**: Supporting the compilation of relevant academic literature, datasets, and tools.

All AI-generated content has been thoroughly **reviewed, revised, and validated** by me to ensure accuracy, academic integrity, and relevance to the project goals. I take full responsibility for the final submission, including all interpretations, methodological choices, and written material.

Generative AI was used strictly as a **creativity and productivity aid**, not as a replacement for original thinking or independent work. The structure, analysis, and conclusions presented in this project reflect my own understanding and critical engagement with the subject matter.

The project in Kaggle notebook: https://www.kaggle.com/code/georgioszachos/nlp-project