

# Project Report

## Text Analytics

### **Political Bias Classification and Bias Analysis**

Mudasir Wazir

## Contents

Executive Summary:.....	3
Problem Definition & Significance: .....	4
Methodology .....	6
Experiments and Results .....	8
Insights and Future Work .....	12
References: .....	13
Appendix: .....	14

## Executive Summary:

This project explores the multifaceted nature of political bias in news articles, moving beyond simple left-versus-right classification to examine how media outlets emotionally frame content and use sensationalist language to influence readers. We designed and trained a multi-task NLP pipeline that includes political bias classification, emotion tagging, and sensationalism scoring as well as generation left leaning article from a right leaning article.

We employed transformer-based models—specifically fine-tuned BERT and DistilBERT variants—coupled with custom Artificial Neural Network (ANN) classifier heads. In addition, we leveraged generative models like GPT-3.5 Turbo for emotion scoring and sensationalism labeling, allowing us to compare conservative token-based approaches with contextual, generative ones.

### Key findings include:

- **Political Bias Classification:** Our best-performing model (BERT-base-cased + custom ANN) showed strong performance in classifying articles as Left or Right, demonstrating that contextual embeddings capture ideological markers effectively.
- **Emotion Analysis:** Right-leaning articles contained significantly higher expressions of anger, fear, and sadness, while Left-leaning content showed marginally higher admiration and amusement. Center articles maintained the most emotionally neutral tone.
- **Style Transfer & Semantic Similarity:** We used GPT-3.5 Turbo to rephrase articles in different ideological styles. Despite changes in tone and sentiment, the semantic similarity between original and generated versions remained high (cosine similarity scores ranging from 0.864 to 0.922), confirming that factual content can be reframed ideologically without altering core information.

Together, these insights suggest that political bias is not just reflected in what is reported, but how it is reported—through emotion and sensationalism. Our work provides a foundation for automated systems that not only detect bias but also help users critically assess the emotional framing and ideological slant in the media they consume.

## Problem Definition & Significance:

In a hyper-connected media ecosystem, political narratives are increasingly shaped not only by factual reporting but by emotional framing, ideological tone, and sensationalistic language. While traditional research has focused on textual bias through lexical choices or source affiliations, relatively little attention has been paid to how emotions and exaggeration contribute to perceived bias. This project seeks to answer:

- How do emotional expressions differ across political orientations in news media?
- Is there a detectable pattern of sentiment polarity (positive, negative, neutral) that aligns with ideological leanings?
- To what extent is sensationalism used by media sources with different political biases?
- Can we build an effective multi-task model to classify political bias while also identifying underlying emotional and sensational tones?

### Significance:

Political bias in journalism can shape public opinion, influence voter behavior, and even alter the perception of objective truth. Therefore, identifying and understanding the underlying emotional and sensationalist cues in politically charged content is essential for promoting media literacy and healthy public discourse.

This research is significant for several reasons:

- **Multi-Dimensional Bias Detection:** By incorporating not only political orientation but also emotional tone, sentiment polarity, and sensationalism levels, our study offers a comprehensive view of media bias. This allows for richer interpretability and more actionable insights than traditional binary classifiers.
- **Emotional Framing & Public Perception:** Emotion plays a crucial role in how people interpret and remember information. Our findings show that Right-leaning content tends to use more emotionally intense language particularly negative emotions like anger and fear which could increase reader engagement while reinforcing ideological divides.
- **Application Potential:** The models developed here can be used in various real-world applications media bias monitors, news summarization tools, and content moderation systems to assess and communicate ideological leanings and rhetorical techniques to users in real time.

In essence, this project contributes to the growing field of computational journalism by offering a data-driven approach to decode not only what but the *how* of political communication in news media.

## Prior Literature Review

Understanding and detecting political bias in news content has become a central concern in the age of algorithm-driven media consumption. Recent research has explored this issue through the lenses of model interpretability, dataset development, and bias propagation in large language models (LLMs). Below, we summarize key contributions from recent literature that informed our project.

A 2023 study titled *“From Pretraining Data to Language Models to Downstream Tasks”* revealed how political biases inherent in training data are learned and perpetuated by language models. These biases not only affect model outputs but also influence fairness in downstream tasks like hate speech detection and misinformation classification. The paper emphasized the need for bias quantification along social and economic axes and demonstrated that bias is not just present but actively propagated by fine-tuned models.

A 2022 scoping review analyzed 154 NLP studies on political polarization, highlighting a methodological skew toward U.S.-based contexts, Twitter datasets, and predictive modeling over explanatory approaches. It called for stronger interdisciplinary collaboration between social science and machine learning, cautioning against inflated results when models ignore contextual and domain-specific factors.

In *“Decoding News Bias”* (Shah et al., 2025), the authors proposed a multi-bias detection framework across seven categories including political, racial, and sensationalist bias. They found that BERT consistently outperformed other models (e.g., RoBERTa, ALBERT, XLNet) in detecting political bias, achieving an F1-score of 0.89. This supports the decision to fine-tune BERT variants in our own project.

Reddy et al. (2019) introduced the Headline Attention Network for detecting bias in Telugu news. Their model used bidirectional LSTMs and a headline-informed attention mechanism, outperforming baseline classifiers. This work underscores the importance of contextual and structural cues especially headlines in understanding bias, even across non-English datasets.

The *Contextualized Bi-Directional Dual Transformer (CBDT)* model (Raza et al.) addressed bias detection at both the content and lexical level. By integrating two complementary transformers, one for holistic bias detection and one for biased term identification, the

authors significantly improved precision and recall. This work illustrates the effectiveness of multi-level contextual models for nuanced bias detection.

Spinde et al. presented the BABE dataset, annotated by trained experts at both word and sentence levels. Their BERT-based model, trained using distant supervision, achieved state-of-the-art performance. Notably, BABE's high annotation quality helped improve generalization across multiple news platforms, highlighting the value of expert-curated datasets in mitigating subjective bias detection.

These studies collectively emphasize that political bias is multifaceted, embedded in word choice, sentence structure, and topic framing and that transformer-based models, particularly BERT, are effective tools for capturing these dimensions. Our work builds upon these insights by implementing and evaluating fine-tuned BERT variants for classifying political bias in English-language news articles.

## Methodology

### Text Preprocessing

The dataset consisted of 17,362 news articles, each labeled as Left (7,803 articles), Center (3,996 articles), or Right (5,563 articles). The data was sourced from the AllSides dataset, which categorizes media bias based on expert and public evaluations.

The input dataset comprises news article texts and corresponding bias labels. We retain the raw textual content without altering linguistic style, as the model needs to learn from implicit sentiment and framing cues often embedded in such language.

Before analysis, the text was cleaned to remove noise such as:

- Photo credits (e.g., "Photo: Reuters")
- Timestamps (e.g., "Updated at 11:10 a.m. ET")
- Irrelevant captions (e.g., "hide caption")
- Special characters and formatting issues

This step ensured that the analysis focused on the core content of the articles rather than metadata or stylistic elements.

Next sections detail how each analysis was performed.

## Bias Classification

The objective of this project is to identify political bias in news articles using transformer-based language models. We follow a supervised learning pipeline, utilizing a labeled dataset of approximately 17,000 articles categorized by bias intensity. The methodology consists of several key steps:

### Model Selection

To evaluate model robustness across different BERT-based architectures, we fine-tune the following pretrained transformer models:

- ***bert-base-uncased***
- ***bert-base-cased***
- ***distilbert-base-uncased (lighter, faster alternative)***

These models are selected to compare performance trade-offs between model size, case sensitivity, and whole-word masking.

### Tokenization and Input Representation

We use the corresponding tokenizer for each BERT variant. Articles are tokenized with padding and truncation to a maximum length of 256 tokens. The resulting input IDs and attention masks are fed into the models along with integer-encoded bias labels.

### Fine-tuning Setup

Each model is fine-tuned using the Hugging Face transformers and PyTorch libraries. The setup includes:

- Loss function: CrossEntropyLoss (in-built in model)
- Optimizer: AdamW with weight decay
- Learning rate: 2e-5
- Scheduler: Linear warmup with no warmup steps
- Batch size: 16 (train), 32 (validation)

## Emotion Analysis

To measure emotional intensity, a random sample of the dataset was processed using a pre-trained emotion classification model (*distilbert-base-uncased-emotion*). Further, we used *GPT 3.5 turbo* to classify emotions as well and compared the two results. The models scored each article on six emotions: Sadness, **Joy, Love, Anger, Fear and Surprise**.

The results were aggregated by political leaning to identify trends.

## Article Text Generation in different political styles

To explore political framing and narrative transformation, we used GPT-3.5 Turbo to generate left-leaning versions of right-leaning articles by using:

- **Model Used:** GPT-3.5 Turbo
- **Input:** Right-leaning news articles
- **Prompt Objective:** Generate a rewritten version of the article that reflects a **left-leaning perspective**, while retaining factual consistency.
- **Token Limit:** Set to **1024 tokens** to manage response length and maintain focus.

To assess the quality and impact of the generated articles, I conducted a **comparative evaluation** using both sentiment and similarity analyses to compare the tone and content similarity between the original and generated text version of the article as well as looking at the text itself manually.

## Experiments and Results

### 1. Bias Classification:

To assess the feasibility of automatic political bias detection in news articles, we conducted a series of classification experiments using BERT-based models. The goal was to classify articles as either **Left**, **Center**, or **Right** based on their textual content.

**Dataset:** The AllSides Media Bias dataset comprises **17,362** news articles labeled by expert editors as **Left (0)**, **Center (1)**, or **Right (2)**.

- **Input:** Raw article text
- **Target:** Political orientation label (0, 1, 2)
- **Label Distribution:** Balanced across classes to ensure fair training.
- 

Each article was tokenized using the corresponding BERT tokenizer, with a maximum sequence length of 256 tokens. We used the [CLS] token's hidden state from BERT as the contextual representation, followed by a custom classification head.

### Model Architecture

Each model used a custom fully connected neural network head:

- Architecture: 768 → 256 → 128 → softmax



- Activation: ReLU
- Regularization: Dropout (rate = 0.3)
- Loss: CrossEntropy
- Optimizer: AdamW (lr=2e-5)
- Scheduler: Linear learning rate decay
- Early Stopping: Patience = 2 based on validation F1

*(For Performance Metrics see appendix)*

## **Observations:**

### 1. High Overall Accuracy and F1:

The BERT-base-cased model achieved a strong overall accuracy of 89.60% and a macro F1-score of 89.60%, indicating its robustness in handling multi-class political bias classification.

### 2. Class-wise F1 Score Trends:

- Left-leaning articles were classified with exceptional accuracy, achieving the highest F1 score of 92.04% and precision of 92.78%.
- Right-leaning articles also showed reliable results, with an F1 score of 89.04% and precision of 90.37%.
- Center articles presented more difficulty, with a comparatively lower F1 score of 85.62% and recall of 87.52%, suggesting some overlap with Left and Right content.

### 3. Center Class Is Most Ambiguous:

According to the confusion matrix, 90 center articles were misclassified as Left, and 30 as Right, indicating the model struggles to distinguish Center from more polar perspectives—likely due to stylistic and rhetorical similarities.

### 4. Balanced Trade-off Between Precision and Recall:

Across all three classes, both precision and recall values are consistently strong (mostly above 87%), reflecting the model's balanced performance without overfitting to any single class.

## 2. Emotion Scoring by Political Leaning:

A 40% random sample of the dataset was analyzed to assess emotional intensity. The results were grouped by political bias (Left, Center, Right) to determine differences in emotional expression.

label_name	anger	fear	joy	love	sadness	surprise
Center	0.555122	0.134222	0.215239	0.006213	0.087113	0.002091
Left	0.562111	0.129473	0.210707	0.007226	0.085093	0.005392
Right	0.553004	0.133339	0.222702	0.006010	0.081929	0.003016

**Table 1. Results from Distilbert-base-uncased-emotion**

*\*Scores are distributed across all emotions such that the total sum per row equals 1 (probabilistic distribution)*

label_name	anger	fear	joy	love	sadness	surprise
Center	0.325500	0.443750	0.098500	0.015250	0.684750	0.303500
Left	0.352051	0.424231	0.096923	0.016026	0.698333	0.314487
Right	0.385971	0.414568	0.089029	0.012590	0.698201	0.307734

**Table 2. Results from GPT-3.5 Turbo**

*\*Each emotion's score is independent and scaled between 0–1*

- **DistilBERT** generally outputs **lower scores** for emotions across the board, suggesting a more conservative intensity scale.
- **GPT-3.5 Turbo** scores negative emotions (fear, sadness, anger) much higher, indicating it interprets content as more emotionally charged.

Overall, while both models agree on the **emotional valence (e.g., more anger and sadness in right-leaning articles)**, GPT suggests a more emotionally intense media landscape.

This suggests that the degree of **sensationalism** in political reporting **may correlate with ideological leaning**, with right-leaning media more likely to employ sensationalistic language or framing

### 3. Article Generation in Different Political Styles:

This experiment provides a structured way to analyze how political bias can be reframed.

Example 1:	Example 2:
<ul style="list-style-type: none"><li>• <b>Right Article Sentiment Score:</b> -0.653</li><li>• <b>Left Article Sentiment Score:</b> -0.557</li><li>• <b>Cosine Similarity:</b> 0.922</li></ul>	<ul style="list-style-type: none"><li>• <b>Right Article Sentiment Score:</b> -0.624</li><li>• <b>Left Article Sentiment Score:</b> -0.552</li><li>• <b>Cosine Similarity:</b> 0.864</li></ul>

**Table 3. Comparative evaluation of the original and generated article text for two example articles.**

- **Sentiment Scores:** In both examples, the **right-leaning articles** display slightly more **negative sentiment** compared to the **left-leaning** versions.
- **Cosine Similarity:** The cosine similarity scores indicate a high degree of **semantic similarity** between the original and generated articles. The values suggest that, despite stylistic and ideological shifts, the generated left-leaning articles maintain a strong alignment with the factual content of the original right-leaning articles.

*These results suggest that while ideological reframing through **GPT-3.5 Turbo** can successfully alter the **tone** of an article, the **underlying sentiment** can shift differently depending on the subject matter. The high **semantic similarity** indicates that the model preserves the core information even when reinterpreting the narrative.*

## User Interface

### News Article Political Bias Identifier

Enter the URL of a news article:

<https://www.nytimes.com/2025/05/05/world/europe/israel-buildup-soldiers-hamas-gaza.html>

Predict Bias

Article Extracted:

Mr. Netanyahu's opponents quickly criticized the expanded military campaign, saying it would endanger the remaining hostages' lives and would not fundamentally change the dynamic that 18 months of war has wrought. The prime minister's critics — both inside and outside the country — have urged him to bring to an end a conflict that began when Hamas killed about 1,200 people in Israel on Oct. 7, 2023, and took some 250 hostage.

Israeli officials said the offensive would start slowly in anticipati...

Predicted Bias: Right

Finally, a user interface was developed where the user can input an article url to check the bias of that article. It uses a fine-tuned BERT-based neural network classifier,

wrapped in a Streamlit web interface for user interaction. **Article Extraction** was done using the *newspaper3k* library that downloads and parses the article, extracting the main text content.

## Insights and Future Work

The analysis reveals consistent patterns, with subtle nuances depending on the specific topic.

- **Political bias classification** reveals how language patterns and framing choices align with ideological leanings, enabling detection of subtle partisan cues in news content.
- **Emotion:** Negative emotions like **anger** and **sadness** are more pronounced in **right-leaning** content, while **fear** is more prominent in **center-leaning** articles.
- **Sensationalism:** **Right-leaning** articles were found to be more sensationalistic, while **center-leaning** articles showed the least sensationalism.

Further, we find that ideological shifts can be achieved without significant loss of factual content, though the **emotional tone** is influenced by the model's framing.

***Future work** could expand the dataset, analyze topic-specific biases, or incorporate reader perception studies.*

## Implications

This analysis provides actionable insights into media bias without requiring technical expertise. The findings can guide editorial policies, media literacy programs, or further research into public perception of news sources. Practical implications include:

- **Consumers:** Helps readers recognize emotional manipulation in news.
- **Media Outlets:** Encourages more balanced reporting to reduce polarization.
- **Policymakers:** Highlights the need for media literacy programs.

## References:

- Hong, J., Cho, Y., Jung, J., Han, J., & Thorne, J. (2023). Disentangling structure and style: Political bias detection in news by inducing document hierarchy. *Findings of the Association for Computational Linguistics: EMNLP 2023*, 377, 5035–5050. <https://aclanthology.org/2023.findings-emnlp.377>
- Feng, S., Park, C. Y., Liu, Y., & Tsvetkov, Y. (2023). From pretraining data to language models to downstream tasks: Tracking the trails of political biases leading to unfair NLP models. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2836–2850. [https://aclanthology.org/2023.acl-long.159/:contentReference\[oaicite:5\]](https://aclanthology.org/2023.acl-long.159/:contentReference[oaicite:5])
- Spinde, T., Krieger, J. D., Ruas, T., Mitrović, J., Götz-Hahn, F., Aizawa, A., & Gipp, B. (2022). Exploiting transformer-based multitask learning for the detection of media bias in news articles. *arXiv preprint arXiv:2211.03491*. <https://arxiv.org/abs/2211.03491>
- Liu, Y., Zhang, X. F., Wegsman, D., Beauchamp, N., & Wang, L. (2022). POLITICS: Pretraining with same-story article comparison for ideology prediction and stance detection. *arXiv preprint arXiv:2205.00619*. <https://arxiv.org/abs/2205.00619>
- Raza, S., Bamgbose, O., Chatrath, V., Ghuge, S., Sidyakin, Y., & Muaad, A. Y. (n.d.). *Unlocking bias detection: Leveraging transformer-based models for content analysis*. [Conference or Journal name if available].
- Shah, B. S., Shah, D., & Attar, V. (2025). *Decoding news bias: Multi-label detection with LLMs*. [Conference or Journal name if available].
- Spinde, T., Plank, M., Krieger, J. D., Ruas, T., Gipp, B., & Aizawa, A. (n.d.). *Neural media bias detection using distant supervision with BABE - Bias Annotations by Experts*

## Appendix:

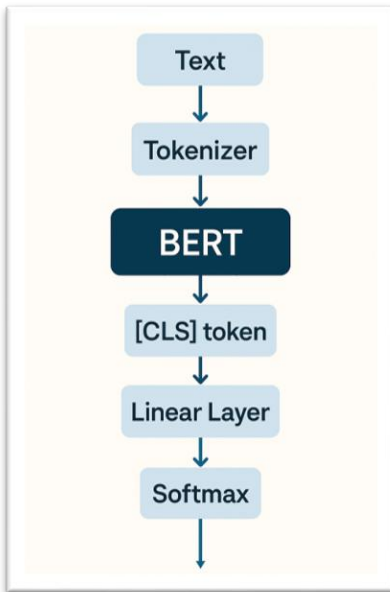


Figure 1. Model architecture for Classification

Metric	Value	Left	Centre	Right
Accuracy	89.65%	-	-	-
Micro F1 Score	89.65%	-	-	-
Micro Precision	89.6%	-	-	-
Micro Recall	89.65%	-	-	-
Precision	-	92.76%	84.38%	89.37%
Recall	-	91.32%	87.53%	88.91%
F1 Score	-	92.04%	85.92%	89.14%

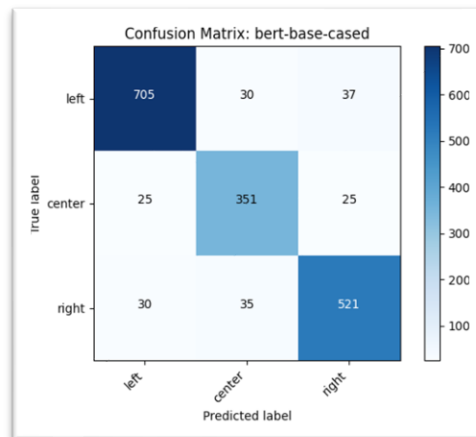


Figure 21. BERT-base-cased performance

sensationalism_label	Not Sensationalist	Sensationalist
label_name		
Center	0.957500	0.042500
Left	0.920513	0.079487
Right	0.811151	0.188849

Figure 32. Sensationalism proportion

In Percentage

```
def generate_left_version(right_text):
    system_prompt = (
        "You are a political speech editor. You will rewrite news articles to reflect a progressive (left-leaning) bias, "
        "emphasizing values such as social justice, government responsibility, environmental protection, and income equality. "
        "Preserve factual accuracy but shift the framing, tone, and emphasis to align with a left-leaning perspective."
    )

    response = openai.chat.completions.create(
        model="gpt-3.5-turbo",
        max_tokens=1024,
        messages=[
            {"role": "system", "content": system_prompt},
            {"role": "user", "content": right_text}
        ]
    )
```

Figure 4. Article Generation Prompt

```
prompt = f"""
You are an expert news analyst.
Determine if the following article is "Sensationalist" or "Not Sensationalist".
Sensationalist articles use exaggerated, emotional, or shocking language.

Article:
---
{truncated}
---

Answer only with "Sensationalist" or "Not Sensationalist"."""

response = openai.chat.completions.create(
    model="gpt-3.5-turbo",
    messages=[{"role": "user", "content": prompt}],
    temperature=0.0,
    max_tokens=10
)
```

Figure 5. Emotion scoring prmt using gpt 3.5 turbo.Emotion scoring prompt using gpt 3.5 turbo.

**Original Right-Leaning Article:** the centers for disease control and prevention (cdc) released a report friday detailing a recent covid-19 outbreak in provincetown, massachusetts, which was cited by the agency as a contributing factor to its recent change in mask guidance. over the july 4 weekend, gatherings in provincetown contributed to at least 900 cases, and roughly 75% of those who tested positive in the cluster were vaccinated. the provincetown report concludes that spread among highly-vaccinated populations does occur and suggests that authorities reimplement mask mandates and other mitigation measures, even for the vaccinated. the provincetown report was a major warning sign for cdc officials, who earlier this week reversed course and once again recommended that vaccinated americans wear face coverings in some indoor settings. a leaked internal document from the cdc included preliminary data from the provincetown outbreak in its summary of information on breakthrough cases of the delta variant. the vaccines are working. of the 900 cases related to the provincetown cluster, there have been no deaths, 7 hospitalizations, and the symptoms are largely mild. our positivity peaked at 15% and was only 4.8% yesterday. the outbreak is contained and provincetown is safe. — alex morse (@alexbmorse) cdc director rochelle walensky said the finding in massachusetts was “concerning and was a pivotal discovery leading to cdc’s updated mask recommendation” in a statement released friday. “high viral loads suggest an increased risk of transmission and raised concern that, unlike with other variants, vaccinated people infected with delta can transmit the virus.” despite the apparently substantial spread among vaccinated individuals that occurred in provincetown, the data clearly indicates that the vaccines work at preventing serious illness and death in nearly all cases. of the more than 900 cases tied to the cluster, only seven required hospitalization, and there have been zero reported deaths nearly a month after the spawn of the outbreak. (related: previously redacted email between peter daszak, anthony fauci suddenly sees light of day) but the high titers is what’s causing concern that even vaccinated can transmit delta, even if they don’t become

**Generated Left-Leaning Article:** The Centers for Disease Control and Prevention (CDC) recently published a report highlighting a COVID-19 outbreak in Provincetown, Massachusetts, which played a significant role in the agency's decision to update its mask guidance. The outbreak, stemming from gatherings over the July 4 weekend, led to around 900 cases, with approximately 75% of those infected being vaccinated individuals. This incident underscored the reality that transmission can occur even among highly vaccinated populations, prompting a call for reinstating mask mandates and other protective measures, irrespective of vaccination status. The Provincetown report served as a critical wakeup call for CDC officials, who swiftly responded by recommending that vaccinated individuals wear masks in certain indoor settings. CDC Director Rochelle Walensky emphasized the importance of the findings from Massachusetts, pointing to the concerning levels of viral load associated with the Delta variant and the potential for vaccinated individuals to transmit the virus. ....

*Figure 6. Article Generation Example.*